

斯蒂格利茨经济学文集

第三卷

中国金融出版社

斯蒂格利茨经济学文集

Selected Works on Economics of Joseph E. Stiglitz

第一卷
信息经济学：基本原理
Vol.1 Economics of Information: Basic Principles

第二卷
信息经济学：应用
Vol.2 Economics of Information: Applications

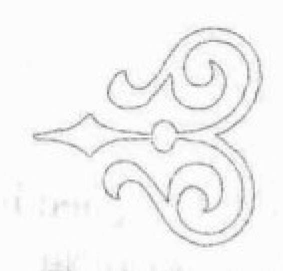
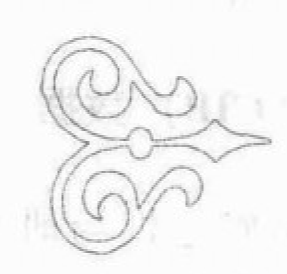
第三卷
微观经济学：不确定性与研发
Vol.3 Micro Economics, Including Uncertainty, R&D

第四卷
宏观经济学：增长与分配
Vol.4 Macro Economics, Growth and Distribution

第五卷
公共财政
Vol.5 Public Finance

第六卷
发展与发展政策
Vol.6 Development and Development Policy

斯蒂格利茨经济学文集


第三卷


微观经济学：不确定性与研发

Micro Economics, Including Uncertainty, R&D

［美］约瑟夫·斯蒂格利茨　著

纪沫　宛圆渊　李鹏飞　译

中国金融出版社

总 策 划：林铁钢
责任编辑：何　为
责任校对：刘　明
责任印制：裴　刚

图书在版编目（CIP）数据

微观经济学：不确定性与研发（Weiguan Jingjixue：Bu Quedingxing yu Yanfa）/（美）斯蒂格利茨（Stiglitz，J. E.）著；纪沫，宛圆渊，李鹏飞译．—北京：中国金融出版社，2007.7
（斯蒂格利茨经济学文集；3）
书名原文：Micro Economics，Including Uncertainty，R&D
ISBN 978－7－5049－4250－0

Ⅰ．微…　Ⅱ．①斯…②纪…③宛…④李…　Ⅲ．微观经济学—文集　Ⅳ．F016－53

中国版本图书馆 CIP 数据核字（2007）第 111878 号

出版发行　中国金融出版社
社址　北京市广安门外小红庙南里 3 号
市场开发部　（010）63272190，66070804（传真）
网 上 书 店　http://www.chinafph.com
（010）63286832，63365686（传真）
读者服务部　（010）66070833，82672183
邮编　100055
经销　新华书店
印刷　北京汇林印务有限公司
尺寸　155 毫米×235 毫米
印张　27.75
字数　503 千
版次　2007 年 9 月第 1 版
印次　2007 年 9 月第 1 次印刷
定价　680.00 元（1—6 卷）
ISBN 978－7－5049－4250－0/F.3810
如出现印装错误本社负责调换　联系电话（010）63263947

目　　录

第三卷导言……………………………………………………………… 1

风险原则……………………………………………………………………… 8
递增的风险：一个定义………………………………………………………… 8
递增的风险：经济意义 ……………………………………………………… 28
投资者偏好和资产回报的结构与投资组合配置的分离性：对共同基金
　纯理论的贡献 ……………………………………………………………… 46

不确定性下的厂商理论 ………………………………………………… 81
关于公司财务纯理论的几个问题：破产与收购 ………………………… 81
股市投资配置的最优性……………………………………………………… 107
价值最大化与企业的可选择目标…………………………………………… 141
关于公司财务政策的无关性………………………………………………… 156

有风险的市场均衡……………………………………………………… 178
论股票市场均衡的无效性…………………………………………………… 178
帕累托次优贸易……………………………………………………………… 204
资本市场自由化、全球化与国际货币基金组织…………………………… 219

垄断竞争理论……………………………………………………………… 240
垄断竞争与最优产品多样性………………………………………………… 240

创新理论…………………………………………………………………… 260
关于技术变革的新观点……………………………………………………… 260
产业结构与创新活动的本质………………………………………………… 266

不确定性、产业结构与研发速度…………………………………………… 297
技术变迁、沉没成本与竞争………………………………………………… 329
干中学、市场结构与产业政策和贸易政策………………………………… 381

福利经济学………………………………………………………………… 405
福利经济学…………………………………………………………………… 405

第三卷导言

标准的竞争均衡模型之所以不能很好地描述实际运行中的市场经济，最重要的原因也许是不完美信息的存在，这也是第一卷和第二卷的主题。但是，在分析风险和不完全的风险市场、不完全竞争和创新时，标准模型的微观基础还有其他一些问题。

本卷（第三卷）的前半部分主要分析风险问题。第一部分，即风险原则部分的文章，为风险的一般理论奠定了概念性基础：当我们说一种情形比另一种情形更有风险时，这究竟是什么意思？风险变大后究竟会带来哪些后果？这些文章对我们理解以前占主导地位的（像均值—方差分析这样的）特定参数方法的局限性产生了重要的影响。

早期的资产组合分析[1]的最重要的一个研究结论就是“资产组合分离定理”。这个定理表明，在一定的限制性条件下，利用均值—方差分析法可以将投资者的投资决策分为两部分：在最优的风险组合中，应该将部分财富投资于资产组合中，同时也要将部分财富投资于无风险资产（政府发行的国库券）之上。只要所有的投资者都根据收入的均值和方差来对投资组合进行评估，那么他们就会购买相同的风险组合。这就意味着市场只需要提供一份共同基金就可以了。在我和戴维·卡斯（David Cass）合作撰写的“投资者偏好和资产回报的结构与投资组合配置的分离性”一文中，我们问道：这些结论到底是否具有一般性？当标准的均值—方差分析法的假设得不到满足时，研究结论会发生哪些变化？我们的分析结果表明，资产组合分离定理赖以成立的前提条件是非常严格的。这毫不让人感到惊讶。但是，将分析结果一般化之后，就得到了两个引人注目的结论。首先，在有些条件下，能满足市场上所有投资者的需求的共同基金的数目是有限的。其次，我们分析了没有无风险资产时的情形。毕竟，由于通货膨胀率是不确定的，所以没有违

1　马科维茨（Markowitz）、夏普（Sharpe）和托宾（Tobin）的研究成果尤为引人注目，他们也因此而获得了诺贝尔经济学奖。

约风险的政府发行的国库券实际上也有风险。在这一情形下，我们证明了双共同基金定理：所有的投资者都要购买一个由市场提供的两种共同基金组成的投资组合，其中一份共同基金是使方差最小化的基金。当然，如果存在完美的无风险资产，那么双共同基金定理就变成了标准的资产组合分离定理。

第二部分（即不确定性下的厂商理论部分）的文章通过考察接管的作用，评价市场的效率，讨论有风险时企业的目标以及刻画企业的融资政策，分析了存在风险的情况下的企业行为。我于 1972 年研究并发表破产和收购（take – over）的文章，在此之前，标准的理论研究文献并没有对这些问题给予足够的重视。其实这也并不让人感到惊讶：它们忽视风险以及只有存在风险时才会发生的破产。此外，即便是有风险，如果企业知道收益的概率分布，贷款人就不会向企业发放太多贷款，从而致使企业破产。在破产成本较高的情况下，这一点更加明显。但是，破产是能发挥重要作用的——破产的威胁会对企业的许多行为产生影响，其中包括限制贷款数量，以及因此而受到影响的企业成长。破产风险可以解释为什么有些企业支付的利率一定要比其他企业支付的利率高。类似地，传统理论不太关注收购也是一件很自然的事情。传统理论告诉我们，管理并不会产生任何差异——所有企业都只是简单地为了实现股价最大化，所以谁拥有或者管理企业并不重要。但是，实际上收购和被收购的威胁是很重要的。尽管在我之前，绝大部分经济学家都忽视了收购，但是还是有些人提到收购是极端重要的。收购实际上相当于在市场上形成了一种纪律。收购能够迫使企业将其市场价值最大化。如果管理者不能做到这一点，其他人就会替代他们，通过改变经营管理政策来获得资本收益。但是，那些持有这些观点的人，一直都没有详细地分析在实际当中收购是怎样发挥作用的。1972 年发表的这篇论文是比较重要的，因为它开创了对破产和收购的研究。文章的分析表明，在不完美信息和风险市场条件下，不同的潜在的所有者可能会有不同的目标。后来，我和桑福德·格罗斯曼（Sanford Grossman）合作撰写的文章“价值最大化与企业的可选择目标”深入分析了这个观点。

我在 1972 年发表的那篇文章同时还表明，收购机制可能会失灵。[2] 它甚至提出了当存在收购时，是否存在均衡的问题。

遗憾的是，这篇文章提出的其他一些重要议题，在后来的研究中并没有得到足够的重视。我发现，对于收益的概率分布，不同的人有不同的看法，

2　格罗斯曼和哈特（Hart）进一步阐述和发展了这些研究结论。

尤其是当前的所有者可能比外部人更加乐观。分析结果表明，在合同的条款是标准条款的情况下，借款人准备支付给贷款人的预期数量会超过贷款人预期收回的数量。这点差异会对那些愿意负债的企业产生重要影响。

在也发表于 1972 年的“股市投资配置的最优性”这篇论文中，我运用均值—方差模型分析了企业的投资行为。由于均值—方差模型提供了一种衡量企业价值的方法，因此被用来明确界定企业价值最大化的投资水平。它假设不同企业的投资收益并不是完全相关的。这意味着投资者希望分散投资，以及企业的股票的需求曲线是一条向下倾斜的曲线。相应地，它还意味着，如果企业的投资增加一倍，即便其产出在任何状态下都增加一倍，企业的价值也不会提高一倍。实际上，这就是“垄断竞争”（参见下文的阐述）。毫不惊讶地，在这些情形中，市场是没有效率的。

新古典的企业理论的一个引人注目的结论是公司的财务政策不重要，即莫迪格利安尼—米勒定理。在第一卷中，我指出了莫迪格利安尼—米勒的分析有一个根本性缺陷，就是假设信息是完美的。在不完美和不对称信息条件下，公司的财务政策能够传递信息，因此就会影响公司的市场价值。在“关于公司财务政策的无关性”这篇文章中，我将这个结论一般化了。我的分析表明，在比莫迪格利安尼和米勒设定的更加一般化的情形下，此研究结论都能成立。我认为，它实际上是一个一般均衡结论。但是，我同时也指出，当存在破产和（与不完美信息有关的）信贷约束时——就像现实世界中所表现出来的那样——公司的财务政策是重要的。因为这些结论似是而非，所以这些结论其实也并不重要——我认为公司的财务政策是相关的——但是它们提供了一些深刻的见解。如果公司的财务政策是重要的，那肯定是由税收、不完美信息或破产引起的。因此，这篇文章有助于那些对公司财务感兴趣的人认识到公司财务政策能产生重要作用。

接下来的部分（即有风险的市场均衡部分）分析了风险市场不完美的（现实）世界中的市场均衡的效率问题。这一部分的第一篇文章的研究结论表明，只要商品的数目大于 1，那么，即便是在企业价值随规模扩大而同比例增长的特殊情形（不同于前一部分所刻画的垄断竞争模型）下，股票市场配置资源的效率也不是帕累托有效的。在更早的时候，戴梦德（Diamond）的分析显示，如果只有一种商品，（从一个特别而又高度限制性的角度看）股票市场就能有效配置资源。从这篇文章中可以得到一个重要的教训：从高度简化的模型中得出与现实世界有关的结论时要十分小心才行。文章结论背后的直觉是，由于风险市场不完全，所以当企业在某个特定产业中增加投资时，就会

改变不同状态下的相对价格，因此也就会改变收益的概率分布。这就好像用一种风险资产去替代另一种风险资产，而且新资产的风险特性也许会优于（或劣于）其他资产。尽管假设政府不能决定最优的投资水平，但是每个企业仍然忽视了不完美的风险市场对收益的概率分布的影响。

不完美的风险市场对价格和利润的概率分布产生的一般均衡效应，是此部分其他两篇论文讨论的核心。在“帕累托次优贸易”这篇文章中，我和戴维·纽伯利（David Newbery）指出，长期以来经济学家支持自由贸易的观念并不一定正确——实际上自由贸易有可能会使每个人的福利水平都下降。批评自由贸易的人通常都会说，自由贸易会使企业暴露在更大的风险之中。我们这篇文章的分析表明，事实确实是这样。在没有自由贸易时，产出增加之后，价格就会下降。此时，价格体系实际上提供了一种保险。的确，在需求弹性为 1 的条件下，不管产出是多少，收益都是一样的。但是，一旦放开贸易，价格就会变得更加稳定。某国的低产出也许就会被其他国家的高产出所抵消。因此，生产就需要承担更大的风险。企业就会从那些产出高度波动的部门中转移出来，哪怕这些部门的收益很高。

信奉自由市场教条的经济学家长期以来都在推动资本市场自由化——他们走得太远了，甚至想在 1997 年 9 月于香港召开的会议上修改国际货币基金组织的章程，以便授权国际货币基金组织去推动许多国家进行资本市场完全自由化的改革。当时，很少有理论和实证文章去分析资本市场自由化的后果。后来，大量的经验证据表明，资本市场自由化可能会增加风险，但是却不能促进投资和增长。本卷中有一篇重要论文（即“资本市场自由化、全球化与国际货币基金组织”），它阐述了为什么国际货币基金组织的政策错过头了——这篇文章的分析表明，资本市场自由化会带来更大幅度的波动和更低水平的福利。文章的关键性结论与前面分析贸易自由化的论文的结论相似。在每代人年轻时储蓄退休后消费的世代交叠的模型中，如果经济体经受一次使工资上升的生产率冲击，那么工人就会储蓄更多的钱，从而提高下期的产出。实际上，以后各代都分享到了生产率冲击带来的利益。在资本市场自由化之后，更高的储蓄只会导致对外国更多的投资。反过来，（像国际利率的变化这样的）来自国外的冲击，会对经济体内部的增长产生扰动作用，资本市场自由化将国家完全暴露在新的不稳定的环境之中。

第三卷的垄断竞争理论部分和创新理论部分考察了不完全竞争和创新问题。市场经济的一个关键性特征就是能够创造出许多多样化的产品。标准的新古典范式忽略了市场经济这个重要特征。在垄断竞争条件下，企业面临着

向下倾斜的需求曲线，因此并不存在一个先验的理由使得市场是有效率的。此外，要生产更多的产品，平均成本就会升高；因此也没有先验的理由让市场去选择最优的产品多样化水平。我和阿维纳什·迪克西特（Avinash Dixit）的研究（“垄断竞争与最优产品多样性”一文）表明，在某种重要的情形下，市场是有效率的，而且每个企业的产出水平和产品多元化决策都是最优的。不过，更一般的情况是，市场的配置都是无效率的。这篇文章为贸易理论和增长理论的大量文献奠定了微观基础。

标准的新古典理论也基本上不考虑创新问题：事实上，它假设技术是固定不变的。但是，市场经济究竟是怎样激励创新活动的？市场机制配置到创新上的资源是多了，还是少了？

近来，增长理论开始关注*内生增长*。在内生增长情形下，创新的速度是在市场内决定的。不过，内生增长理论在罗默（Roemer）那篇引用率很高的论文发表之前实际上就已经有了一定的发展。谢尔（Shell）、宇泽（Uzawa）、尼尔逊（Nelson）、费尔普斯（Phelps）、阿罗（Arrow）等人在这方面做出了重要贡献。在我早期的研究成果中有一个重要的思想，即（作为创新的基础的）知识实际上是一种特殊类型的信息，[3] 与其他所有信息一样，它也面临着可占用性问题。知识是一种公共物品，所以限制知识的使用就是低效率的。结果就是，在知识的市场和使用这两方面，市场可能都是无效率的。于是，政府干预就会发挥重要作用。在有关福利经济学的论文（本卷最后一篇文章）中，我深入地分析了这个观点。

我早期的一项贡献，也就是我与安东尼·阿奇生（Anthony Atkinson）合作撰写的发表于1969年的论文（“关于技术变革的新观点”）重点分析了这样一个事实，企业倾向于（通过学习或者投入资源到研发活动之上）围绕其正在使用的技术来进行创新——技术变革是“局部性的”。这个简单的想法对于为技术变革和经济增长“建模”有重要的含义。它表明，只是根据劳动或资本增加的进步来描述技术变革可能并不准确。尤其是它意味着历史发生作用的滞后效应的存在。某个导致工资提高和使用劳动节约型技术的小插曲，意味着某些技术会得到改进，而另一些技术则会被遗弃。

其他有关创新的论文重新审视了熊彼特关于竞争与创新的关系的观点。熊彼特认为，有技术竞争的市场上，最典型的特征就是，垄断者的成功都是暂时的。不过，我们的分析表明，（不存在政府干预时）垄断企业有可能维

3 尤其要参见 Stiglitz（1975）。

持其垄断地位。

产业结构会对创新的速度产生影响，不过以前的研究表明，两者之间的关系非常复杂。尤其是，我们认为，产业结构本身可能就是内生的。在“产业结构与创新活动的本质”这篇文章中，我们同时解决了创新速度和产业结构问题。

在20世纪80年代，一种新的教条，即可竞争性市场教条开始逐渐兴起。这种观点认为，即便是自然垄断，也不会对经济产生负面影响。竞争是在争夺进入市场时展开的，而不是在市场上展开的：潜在的竞争实际上成了实际竞争的替代物。价格会降低至平均成本水平，而且市场仍然是有效率的（在零利润—无补贴的约束条件下，由于价格等于平均成本，所以它超过了边际成本。但是，固定成本抵消了高价格带来的收益）。在“技术变革、沉没成本与竞争”这篇文章中，我指出，可竞争性市场教条并不稳健：只要有许多沉没成本，进入就会受到阻碍。即便利润很高，潜在竞争对手的进入也会很困难。真正重要的不是进入之前的利润，而是进入之后的利润。假设存在数量较少的固定不变的沉没成本，并且边际成本不变，竞争会使得价格等于边际成本，所以如果进入者的边际成本等于现有企业的边际成本，那么它就无法收回沉没成本。通常沉没成本都会存在，比如航空公司要想让顾客知道开通的航线，它就需要花钱去做广告。这就是沉没成本。一个引人注目的结论是，（一旦进入就会发生的）事后的竞争越激烈，（进入发生之前的）事前的竞争可能就越弱。

在研发领域，沉没成本尤为重要，因此在研发活动很重要的产业中，市场并不是可竞争的。事实上，我的分析表明，如果一家企业领先竞争对手一点点——领先的优势只要大到让竞争对手明白它不可能超越这家企业——那么该企业就可以继续保持领先地位。竞争可能很难为创新提供激励。

更为一般地，我们的分析结果显示，很难说市场能为创新活动提供充足的资金。正如本卷创新理论部分所阐述的那样，有充分的证据表明，发展中国家有必要对幼稚产业进行保护。有人认为，提供补贴的效果更好，也有人认为应当让企业能够借到贷款为学习活动融资。不过，这些观点都忽视了政府征税能力的局限性（第五卷会有详细的讨论），以及企业借款努力的局限性。当信息不完美时，这些局限性体现得尤为冒险。

本卷的最后一部分考察了福利经济学中更为宽泛的主题：没有政府干预，市场本身是不是有效的？如果是有效的，究竟是哪些因素在发挥作用？新的信息范式推翻了经济学最重要最悠久的一个结论，即福利经济学基本定

理，有时也称为亚当·斯密的看不见的手。福利经济学基本定理认为竞争的市场是（帕累托）有效的。这一部分的论文推翻了上述结论：我们的分析结果表明，只要存在不完美信息或者市场是不完全的——通常情况都是这样的——则在一定的范围内政府干预能使得每个人的境况都变得更好。最后一篇文章分析了为什么是这样的，而且以政府干预能发挥重要作用的具体例子展示了这个一般性的理论。

风险原则

递增的风险：一个定义*

一、绪论

本文试图回答以下问题：在什么时候，随机变量 Y 比另一个随机变量 X 更加易变？

直觉和传统对此至少给出了四种看似合理——并且明显不同——的答案。它们分别是：

1. Y 等于 X 加上噪音

如果我们只是对一个随机变量加上某个不相关的噪音，这个新的随机变量会比原来的随机变量更具风险性。[1] 更加正式的表述是，假设 Y 和 X 有如下关系：

$$Y \underset{d}{=} X + Z \tag{1. i}$$

其中“$\underset{d}{=}$”表示“具有同样的分布”，Z 是一个随机变量，它具有如下性质：

$$E(Z|X) = 0 \text{ 对所有的 } X \text{ 成立}^{2} \tag{1. ii}$$

---

* “Increasing Risk: I. A Definition”, with M. Rothschild, *Journal of Economic Theory*, 2 (3), September 1970, pp. 225-243。本文的研究课题得到了美国国家科学基金会（National Science Foundation）和福特基金会（Ford Foundation）的资助。

1 在本文中，“更加易变”、“更具风险性”、“更不确定”表示同样的意思。

2 戴维·华莱士（David Wallace）建议我们研究这一“更大风险性”的概念。阿瑟·古德伯格（Arthur Goldberger）指出，式（1. ii）比我们在以前版本的论文中给出的“缺乏相关性”条件更强。

也就是说，Y 等于 X 加上一个干扰项（噪音）。如果 X 和 Y 是离散随机变量，条件（1）有另一种自然的解释。假设 X 是一个彩票，收益为 a_i 的概率是 p_i；$\sum p_i = 1$。而 Y 也是一种彩票，收益 b_i 的概率是 p_i，其中 b_i 要么是 a_i 的回报，要么是一个期望收益为 a_i 的彩票。需要注意的是，条件（1）意味着 X 和 Y 具有同样的均值。

2. 每个风险厌恶者在 X、Y 之间更偏好 X

在预期效用最大化理论中，风险厌恶者被定义为一个具有凹形效用函数的人。如果 X 和 Y 具有同样的均值，但是每个风险厌恶者在 X 和 Y 之间更偏好 X，即

$$EU(X) \geqslant EU(Y) \text{ 对所有凹的效用函数都成立} \tag{2}$$

则称 X 比 Y 风险性更小是非常合理的。[3]

3. Y 在双尾的权重比 X 在双尾的权重大

如果 X 和 Y 的密度函数分别是 f 和 g，其中 g 由 f 转换而来，它们具有同样的均值，但是 g 将 f 的接近中心的一些概率权重挪到了双尾，这样称 Y 比 X 更不确定就似乎很合理了。

4. Y 比 X 的方差大

关于风险性或者不确定性的比较通常局限于关于方差的比较，这主要因为长久以来，在统计学理论中都将方差作为衡量分散性的指标。

本文的主要结论是：前三种方法引出了一个关于更大风险性的相同的定义，该定义和第四种方法引出的定义有所不同。我们将如下展示其等同性。在第二部分，我们将证明第三种方法会引出一个以累积分布函数（cumulative distribution function）的差的无穷积分表示的关于递增的不确定性的描述。在第三部分中，我们将证明这一无穷积分会产生在分布函数集上的偏序（partial ordering），这和前两种方法产生的偏序相同。

在本文第四部分中，我们将证明，这一关于递增的风险性的概念并不等同于将 X 的风险性和 X 的方差相提并论。这意味着相对于标准方法而言，由我们的概念引出的关于递增风险的定义更好。

当然，我们不可能证明一条定义比另一条定义好。但是，这一事实并不意味着可以持有不可知论或者可以停止评判。尽管对我们而言，我们的定义

---

3 可能有人会认为，我们的讨论应该限于凹的增函数。增加这一限制毫无益处，反而会破坏一些结论的对称性。比如，由于 $U(X) = X$ 和 $U(X) = -X$ 都是凹函数，条件（2）意味着 X 和 Y 具有同样的均值。

与方差的定义相比，更符合递增的风险的自然含义。但定义的选择既需要考虑其一致性，也要考虑其有用性。正如托宾（Tobin）指出的，要批评均值方差的方法，“不仅需要证明该方法是依赖于限制性假设的，还需要给出一个更加宽泛且难以攻击的方法，并且说明该方法是如何获得经济学家们感兴趣的比较静态结果的［参考文献8］”。在本文的结尾，我们将展示该定义如何被应用于经济学和统计学问题。

在本文展开之前，我们需要给出关于一些标记的惯例。在本文中，X 和 Y 表示分别具有累积分布函数 F 和 G 的随机变量。当存在 F 和 G 的密度函数的时候，我们将其分别记为 f 和 g。总体上，我们将保持惯例，认为 F 的风险性比 G 小。

目前，我们的结果仅对增值点（points of increase）在有界区间内的累积分布函数有效，为简便起见，我们将此区间设为［0，1］，也就是 $F(0)=G(0)=0$，且 $F(1)=G(1)=1$。要将此结果扩展到定义在整个实数轴上的累积分布函数，这一问题还有待解决，并且这一问题的解决需要先解出一系列没有任何经济意义的微妙的收敛问题。$H(x, z)$ 是定义在［0，1］×［-1，1］（即［0，1］和［-1，1］的笛卡儿积）上的随机变量 X 和 Y 的联合分布函数。令 S 表示 G 和 F 的差别，T 表示 S 的无穷积分。也就是说，$S(x)=G(x)-F(x)$, $T(y)=\int_0^y S(x)dx$。

二、积分条件

一个随机变量比另一个随机变量在尾部上的权重大意味着什么？在本部分，我们将对此给出一个几何上的定义（本部分第1点、第2点）。“更高的风险”这个概念应当是可以传递（transitive）的。要考察这一要求的后果，就引发了一个更宽泛的定义。尽管这个定义不那么直观，但是更便于分析（第3点、第4点）。

1. 均值保留展型：密度

$s(x)$ 是如下定义的阶梯函数

$$s(x)=\begin{cases}\alpha \geqslant 0 \text{ 对于 } a<x<a+t \\ -\alpha \leqslant 0 \text{ 对于 } a+d<x<a+d+t \\ -\beta \leqslant 0 \text{ 对于 } b<x<b+t \\ \beta \geqslant 0 \text{ 对于 } b+e<x<b+e+t \\ 0 \text{ 其他}\end{cases} \tag{3.i}$$

其中

$$0\leqslant a\leqslant a+t\leqslant a+d\leqslant a+d+t \leqslant b\leqslant b+t\leqslant b+e\leqslant b+e+t\leqslant 1 \tag{3. ii}$$

并且

$$\beta e=\alpha d \tag{3. iii}$$

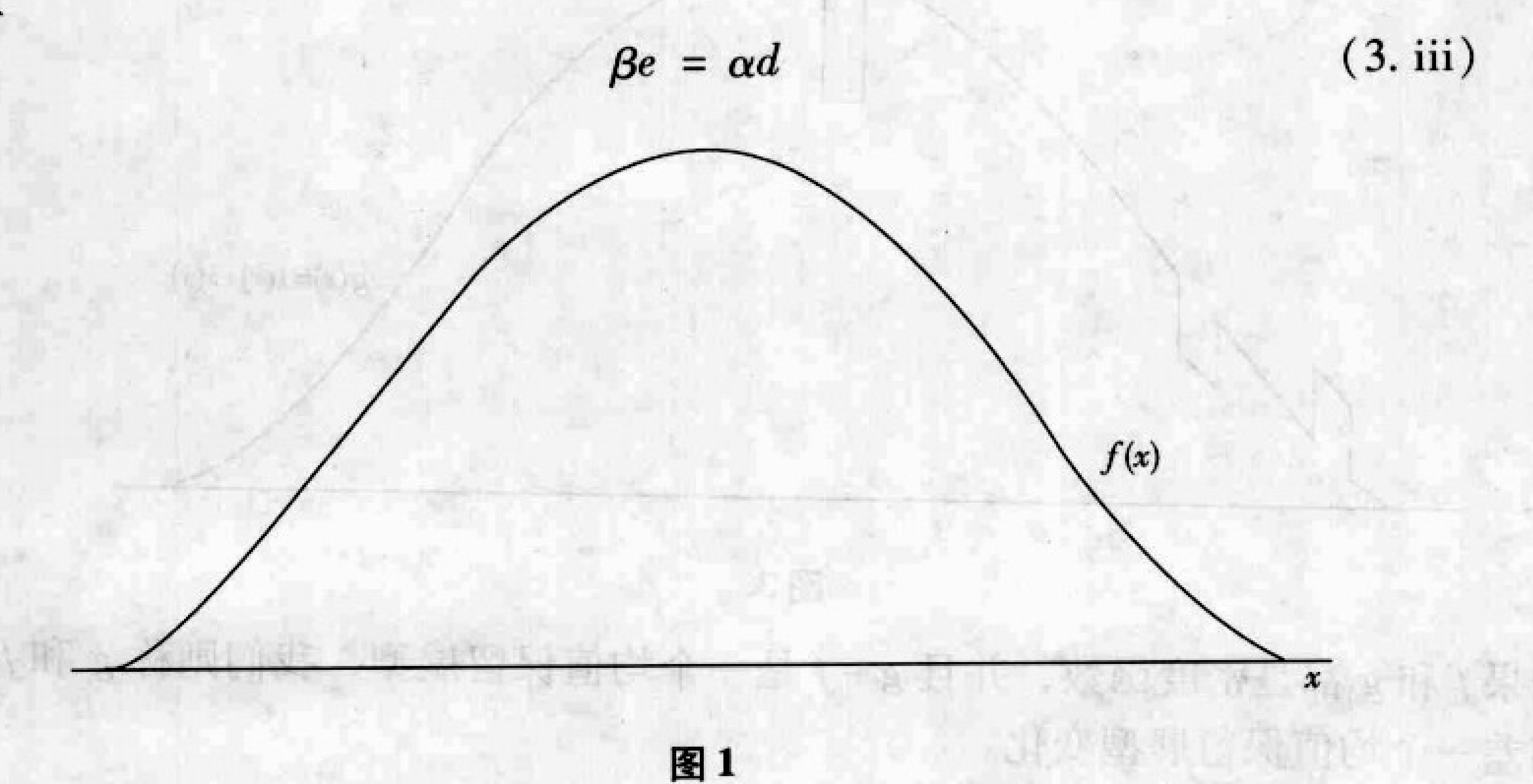


图 1

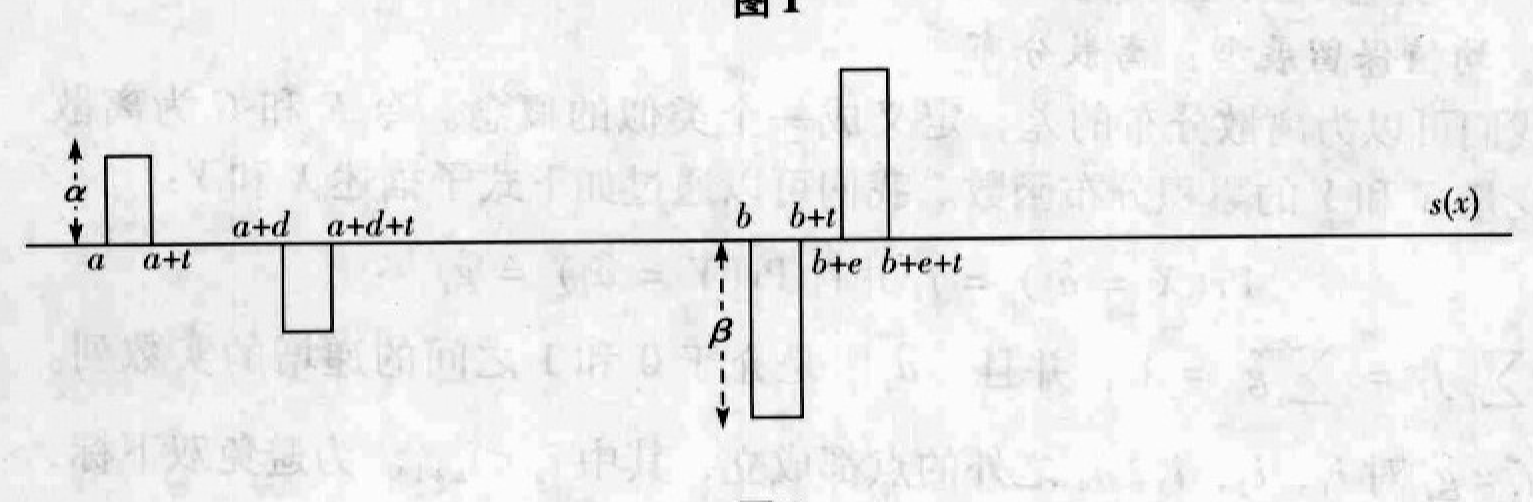


图 2

该函数如图 2 所示。容易验证，$\int_0^1 s(x)dx=\int_0^1 xs(x)dx=0$。因此，如果$f$是一个密度函数，并且 $g=f+s$，则 $\int_0^1 g(x)dx=\int_0^1 f(x)dx+\int_0^1 s(x)dx=1$ 并且 $\int_0^1 xg(x)dx=\int_0^1 x(f(x)+s(x))dx=\int_0^1 xf(x)dx$。由此可得，如果 $g(x)\geqslant 0$ 对所有 x 都成立，g 就是一个和 f 具有相同均值的密度函数。[4] 对 f 添加一个类似于 s 的函数可以将概率权重从中心挪到双尾。参见图 1 和图 3。我们把满足条件(3) 的函数称为均值保留展型（Mean Preserving Spreads，MPS）。

4 也就是说，如果在 $a+d<x<a+d+t$ 上 $f(x)\geqslant\alpha$，在 $b<x<b+t$ 上 $f(x)\geqslant\beta$。

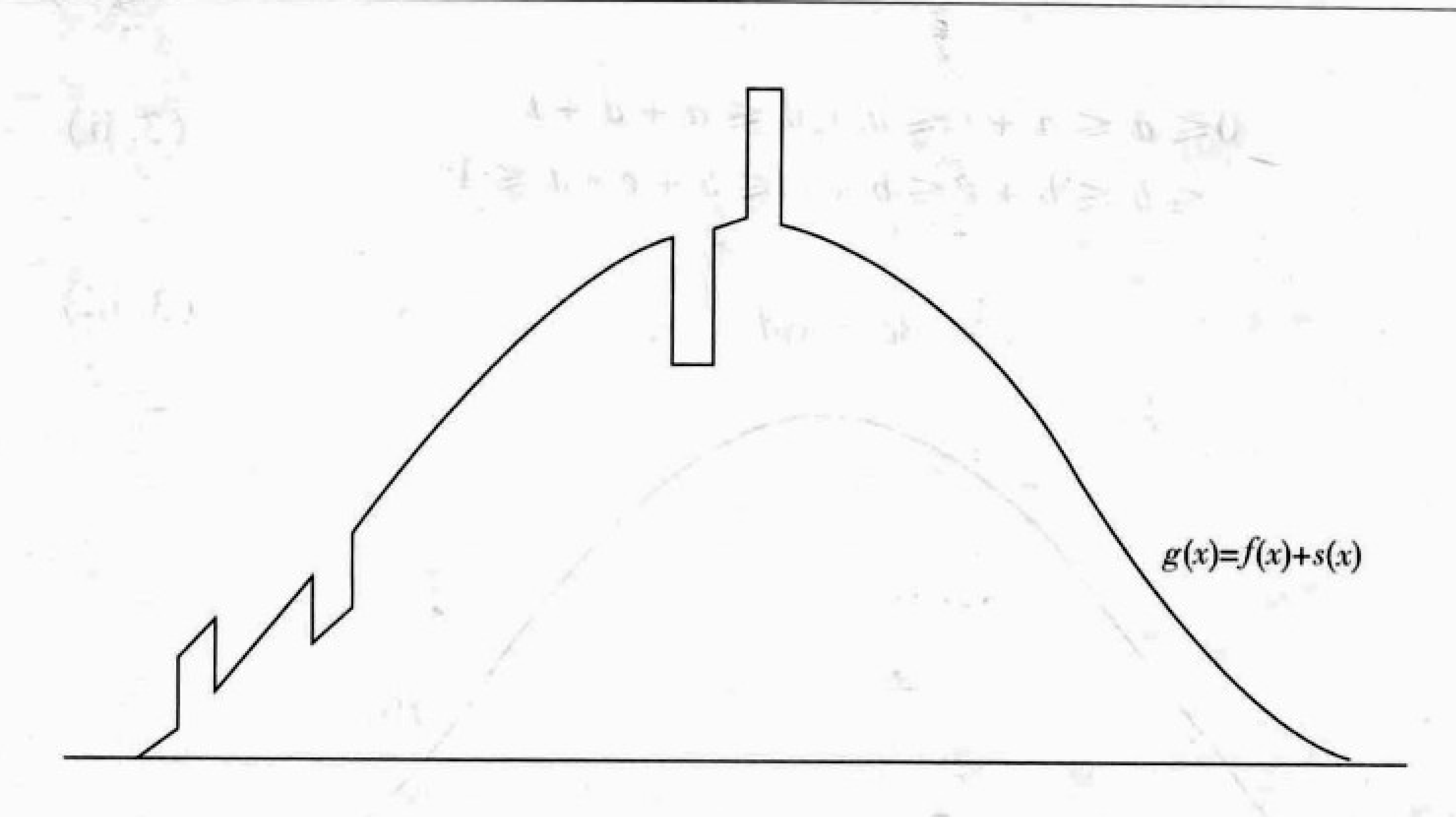


图3

如果f和g都是密度函数，并且$g-f$是一个均值保留展型，我们则称g和f相差一个均值保留展型变化。

2. 均值保留展型：离散分布

我们可以为离散分布的差，定义成一个类似的概念。令F和G为离散随机变量X和Y的累积分布函数。我们可以通过如下式子描述X和Y：

$$\Pr(X=\hat{a}_i)=\hat{f}_i \text{ 并且 } \Pr(Y=\hat{a}_i)=\hat{g}_i$$

其中$\sum_i \hat{f}_i=\sum_i \hat{g}_i=1$，并且$\{\hat{a}_i\}$是介于0和1之间的递增的实数列。假设$\hat{f}_i=\hat{g}_i$对$i_1$，$i_2$，$i_3$，$i_4$之外的点都成立，其中$i_k<i_{k+1}$。为避免双下标，我们令$a_k=\hat{a}_{i_k}$，$f_k=\hat{f}_{i_k}$，且$g_k=\hat{g}_{i_k}$，并且定义

$$\gamma_k=g_k-f_k$$

那么，如果

$$\gamma_1=-\gamma_2\geqslant 0 \quad \text{并且} \quad \gamma_4=-\gamma_3\geqslant 0 \tag{4.i}$$

则Y在尾部的权重比X在尾部的权重大，并且如果

$$\sum_{k=1}^{4} a_k\gamma_k=0 \tag{4.ii}$$

则X和Y的均值相同。参见图4。如果离散随机变量X和Y对这四个点以外的所有点赋予的权重相同，并且二者的差满足条件（4），我们则称Y和X相差一个均值保留展型变化。

3. 积分条件

如果密度函数g和f相差一个均值保留展型变化，s，其相应的累积分

布函数 G 和 F 的差就是 s 的无穷积分。也就是说，$s=g-f$ 意味着 $S=G-F$，其中 $S(x)=\int_0^x s(u)du$。S（如图5所示）具有一些有趣的性质。其中最后两条性质（下面的等式（6）和等式（7））在本文中有重要作用，我们将此作为积分条件。第一，$S(0)=S(1)=0$。第二，存在一个满足下列条件的 z：

$$S(x)\geqslant 0 \text{ 如果 } x\leqslant z \quad \text{并且} \quad S(x)\leqslant 0 \text{ 如果 } x>z \tag{5}$$

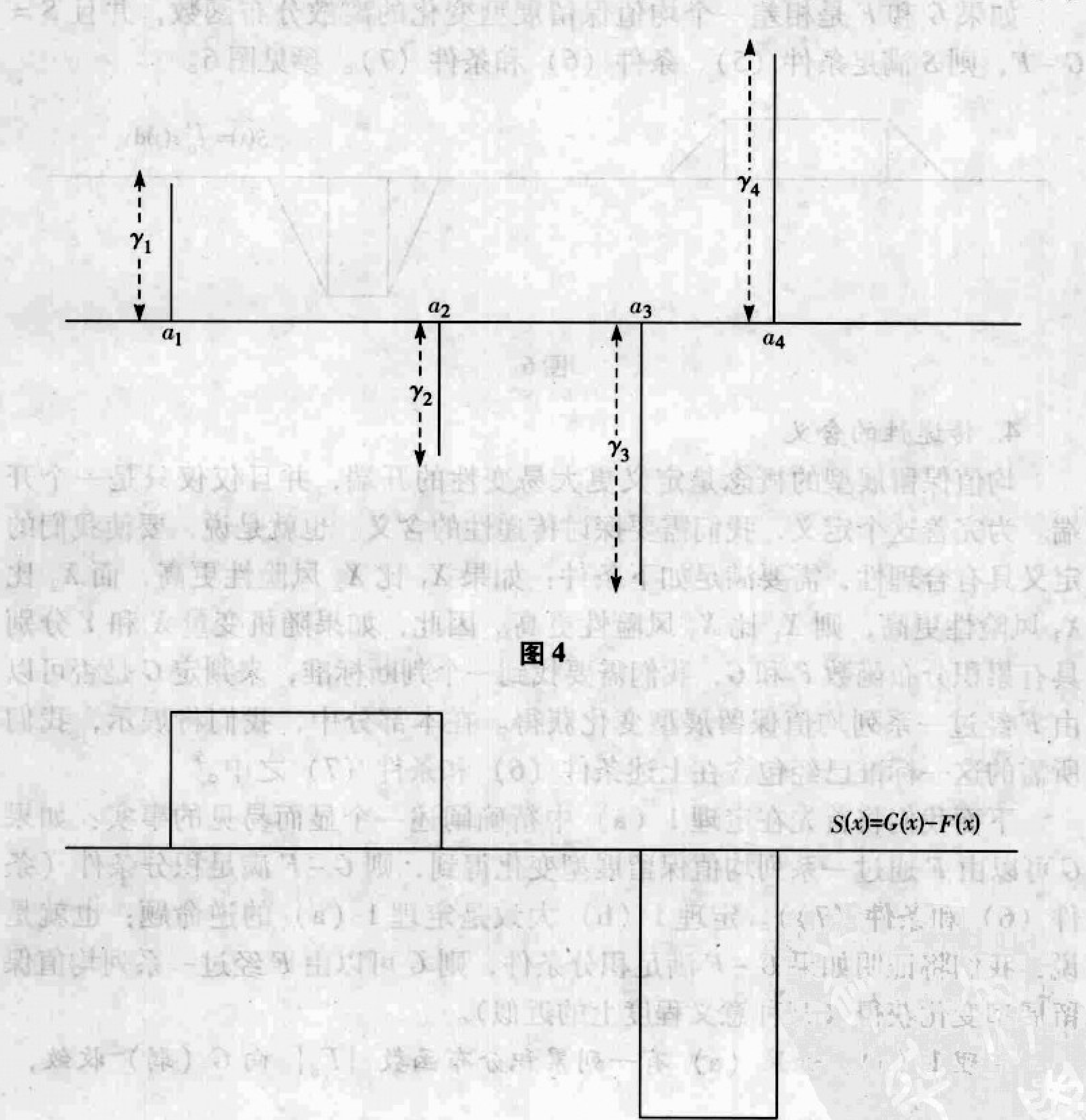


图4

图5

第三，如果 $T(y) = \int_o^y S(x)dx$，则

$$T(1) = 0 \tag{6}$$

这是因为 $T(1) = \int_0^1 S(x)dx = xS(x)]_0^1 - \int_0^1 xs(x)dx = 0$。

最后，条件（5）和条件（6）在一起意味着

$$T(y) \geqslant 0, 0 \leqslant y < 1 \tag{7}$$

如果 G 和 F 是相差一个均值保留展型变化的离散分布函数，并且 $S = G - F$，则 S 满足条件（5）、条件（6）和条件（7）。参见图6。

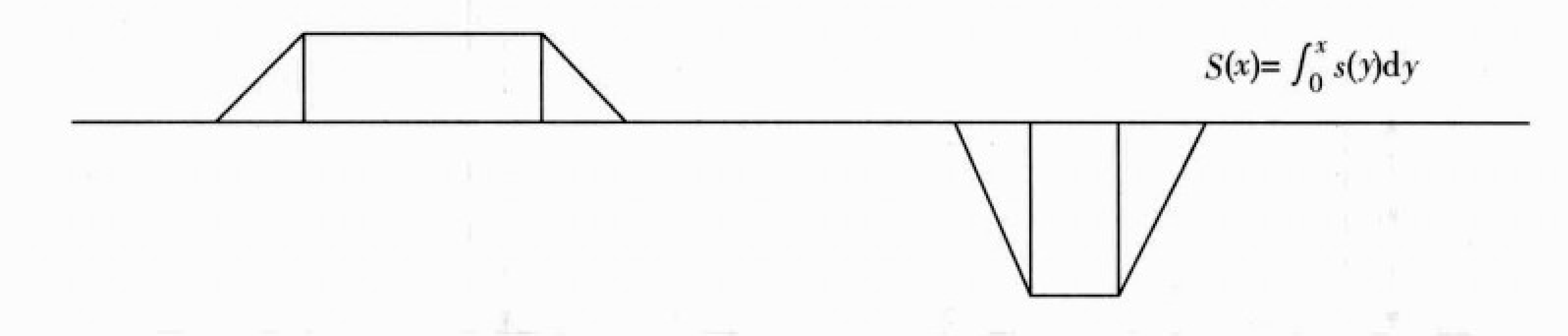


图6

4. 传递性的含义

均值保留展型的概念是定义更大易变性的开端，并且仅仅只是一个开端。为完善这个定义，我们需要探讨传递性的含义。也就是说，要使我们的定义具有合理性，需要满足如下条件：如果 X_1 比 X_2 风险性更高，而 X_2 比 X_3 风险性更高，则 X_1 比 X_3 风险性更高。因此，如果随机变量 X 和 Y 分别具有累积分布函数 F 和 G，我们需要找到一个判断标准，来判定 G 是否可以由 F 经过一系列均值保留展型变化获得。在本部分中，我们将展示，我们所需的这一标准已经包含在上述条件（6）和条件（7）之中。[5]

下面我们将首先在定理1（a）中精确阐述一个显而易见的事实：如果 G 可以由 F 通过一系列均值保留展型变化得到，则 $G-F$ 满足积分条件（条件（6）和条件（7））。定理1（b）大致是定理1（a）的逆命题：也就是说，我们将证明如果 $G-F$ 满足积分条件，则 G 可以由 F 经过一系列均值保留展型变化获得（某种意义程度上的近似）。

定理1（a）：如果（a）有一列累积分布函数 $\{F_n\}$ 向 G（弱）收敛，

5 条件（5）不能包含在该标准中，因为我们可以很容易地构建一对相差两个均值保留展型的累积分布函数，但是其差不能满足条件（5）。

（记为 $F_n \to G$）[6] 并且（b）F_n 和 F_{n-1} 相差一个均值保留展型变化（这意味着 $F_n = F_{n-1} + S_n = F_0 + \sum_{i=1}^{n} S_i$，其中 $F_0 \equiv F$，并且每个 S_i 都满足条件（6）和条件（7）），则 $G = F + \sum_{i=1}^{\infty} S_i = F + S$ 并且 S 满足条件（6）和条件（7）。

其证明显而易见，此处省略。

定理 1（b）：如果 $G - F$ 满足积分条件（6）和条件（7），则存在一列 F_n 和 G_n，$F_n \to F$，$G_n \to G$ 满足：对于每个 n，G_n 可以由 F_n 经过有限的均值保留展型变化获得。

该定理的证明是以下两个引理的直接结果：第一个引理证明了定理对具有有限个阶梯的阶梯函数成立；第二个引理说明 F 和 G 可以通过满足积分条件的阶梯函数无限近似。

引理 1：如果 X 和 Y 是离散随机变量，其累积分布函数分别是具有有限个增值点的 F 和 G，并且如果 $S = G - F$ 满足条件（6）和条件（7），则存在一列累积分布函数 F_0，…，F_n 满足 $F_0 = F$，$F_n = G$，并且 F_i 和 F_{i-1} 只相差一个均值保留展型变化。

证明：S 是一个具有有限个阶梯的阶梯函数。令 $I_1 = (a_1, a_2)$ 是 S 的第一个正的阶梯。如果 I_1 不存在，$S(x) \equiv 0$ 意味着 $F = G$，那么此定理明显正确。令 $I_2 = (a_3, a_4)$ 是 S 的第一个负的阶梯。通过式（7），$a_2 < a_3$。令 γ_1 为 $S(x)$ 在 I_1 上的值，$-\gamma_2$ 是 $S(x)$ 在 I_2 上的值。

或者

$$\gamma_1(a_2 - a_1) \geqslant \gamma_2(a_4 - a_3) \tag{8}$$

或者

$$\gamma_1(a_2 - a_1) < \gamma_2(a_4 - a_3) \tag{9}$$

如果式（8）成立，令 $\hat{a}_4 = a_4$。在 $a_1 < \hat{a}_2 \leqslant a_2$ 区间存在一个 $\hat{a}_2$ 满足

$$\gamma_1(\hat{a}_2 - a_1) = \gamma_2(\hat{a}_4 - a_3) \tag{10}$$

如果式（9）成立，令 $\hat{a}_2 = a_2$；则在 $a_3 < \hat{a}_4 < a_4$ 区间存在一个 $\hat{a}_4$ 使式（10）成立。如下定义 $S_1(x)$：

$$S_1(x) = \begin{cases} \gamma_1 & \text{对于 } a_1 < x < \hat{a}_2 \\ -\gamma_2 & \text{对于 } a_3 < x < \hat{a}_4 \\ 0 & \text{其他} \end{cases}$$

---

6 使 $E(u) = \int_0^1 u(x)dG(x)$ 和 $E_n(u) = \int_0^1 u(x)dF_n(x)$。那么 $F_n \to G$ 当且仅当 $E_n(u) \to E(u)$，对于所有连续的在 $[0,1]$ 上的 u。参见［参考文献 3］。

如果 $F_1 = F_0 + S_1$，则 F_1 和 F 相差一个均值保留展型变化，并且 $S^{(1)} = G - F_1$ 满足条件（6）和条件（7）。

我们采用这个技术来构建 S_2（通过 $S^{(1)}$ 变化而来），并且定义 F_2 为 $F_2 = F_1 + S_2$。由于 S 是一个具有有限阶梯的阶梯函数，这一过程在有限的重复之后就会终结。

引理2：令 F 和 G 是定义在 $[0, 1]$ 上的累积分布函数。令 $T(y) = \int_0^y (G(x) - F(x))dx$。如果

$$T(y) \geqslant 0,\ 0 \leqslant y \leqslant 1 \tag{6}$$

并且

$$T(1) = 0 \tag{7}$$

那么，对于每个 n，都存在具有有限增值点的 F_n 和 G_n（离散随机变量的累积分布函数），满足：如果

$$\| F_n - F \| = \int_0^1 | F_n(x) - F(x) | dx$$

并且

$$\| G_n - G \| = \int_0^1 | G_n(x) - G(x) | dx$$

则[7]

$$\| F_n - F \| + \| G_n - G \| \leqslant \frac{4}{n} \tag{11}$$

并且，如果 $T_n(y) = \int_0^y (G_n(x) - F_n(x))dx$，则

$$T_n(y) \geqslant 0 \tag{12}$$

并且

$$T_n(1) = 0 \tag{13}$$

证明：我们通过对固定的 n 构建 F_n 和 G_n 来证明这条引理。对于 $i = 1, \cdots, n$，令 $I_i = ((i-1)/n, i/n)$。令 $\bar{f}_i = F(i/n)$，并且通过 $\bar{F}_n(x) = \bar{f}_i$（对于所有 $x \in I_i$ 都成立）定义 $\bar{F}_n$（参见图7）。由于 F 是单调变化的，$\bar{F}_n(x) \geqslant F(x)$。同样由于单调性，$\| \bar{F}_n - F \| \leqslant 1/n$。如果 $\hat{F}_n(x)$ 是一个在每个 I_i 区间为常数的阶梯函数，并且满足对于 $x \in I_i$，$\hat{F}_n(x) \in F(I_i)$，则

7 条件（11）意味着弱收敛［参考文献3，p.243］。

$\| \hat{F}_n - \bar{F}_n \| \leqslant 1/n$，并且

$$\| \hat{F}_n - F \| \leqslant \| \hat{F}_n - \bar{F}^n \| + \| \bar{F}_n - F \| \leqslant \frac{2}{n}$$

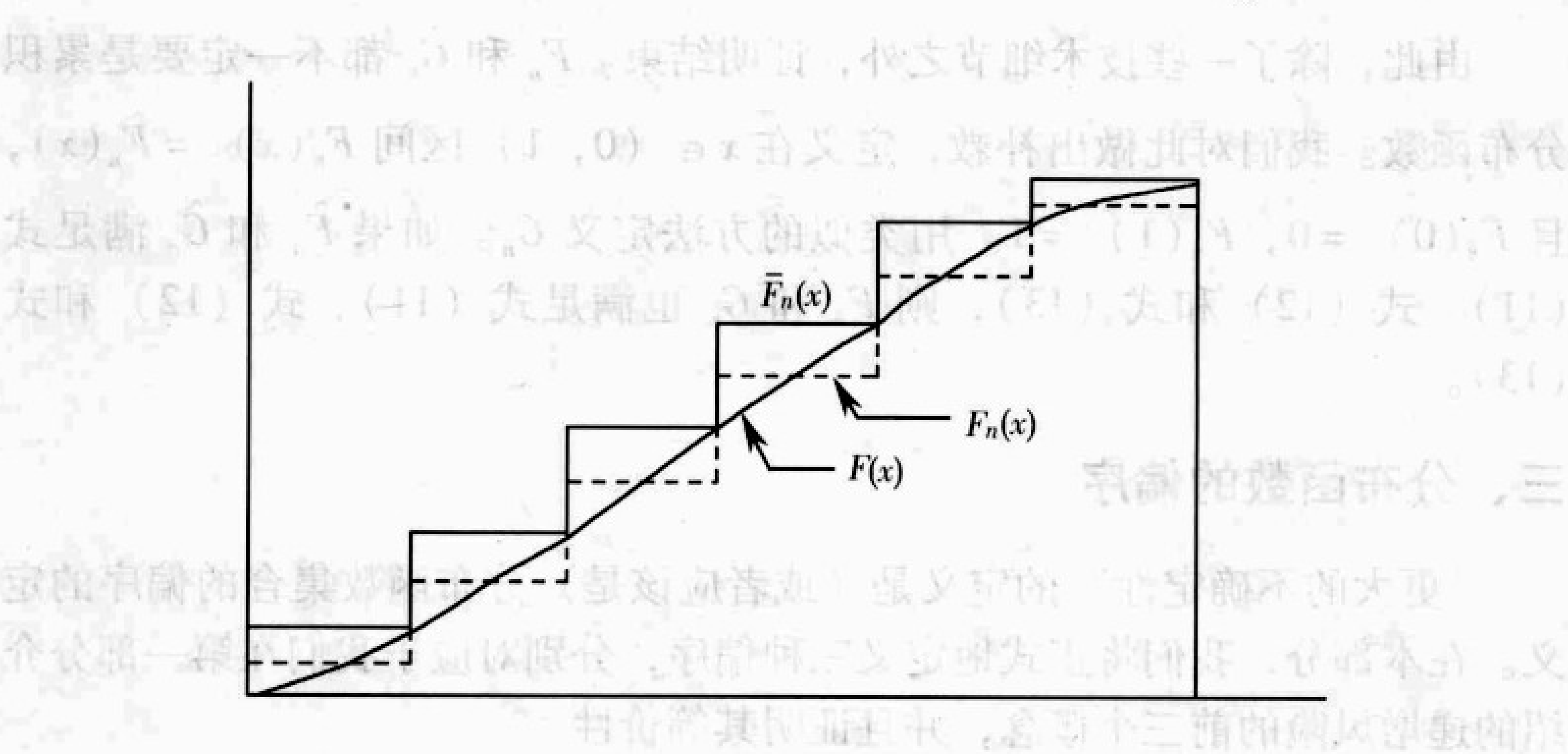


图 7

类似地，如果 $\hat{G}_n(x)$ 是一个阶梯函数，其满足：对于 $x \in I_i$，则$\hat{G}_n(x) \in G(I_i)$，则 $\| \hat{G}_n - G \| \leqslant 2/n$。

对于每一个 i，都存在 $f_i \in F(I_i)$ 和 $g_i \in G(I_i)$，使得 $(g_i - f_i)/n = \int_{I_i}(G(x) - F(x))dx$。对于$x \in I_i$，令 $\hat{F}_n(x) = f_i$，$\hat{G}_n(x) = g_i$。现在我们要证明 $\hat{F}_n$ 和 $\hat{G}_n$ 满足式（11）、式（12）和式（13）。我们已经证明了式（11）是满足的。我们观察到

$$\begin{aligned}\hat{T}_n(1) &= \int_0^1 (\hat{G}_n(x) - \hat{F}_n(x))dx \\ &= \sum_{i=1}^n \int_{I_i}(\hat{G}_n(x) - \hat{F}_n(x))dx \\ &= \sum_{i=1}^n \frac{g_i - f_i}{n} = \sum_{i=1}^n \int_{I_i}(G(x) - F(x))dx \\ &= \int_0^1 (G(x) - F(x))dx = T(1) = 0\end{aligned}$$

因此，式（13）是满足的。我们还需证明的是 $\hat{T}_n(y) \geqslant 0$。如果对于$j = 0, 1, \cdots, n$，$y = j/n$，则 $\hat{T}_n(y) = T(j/n) \geqslant 0$。因此我们只需要检验 $y = j/n + \alpha$，$0 < \alpha < 1/n$ 的情况。因此 $\hat{T}_n(x) = T(j/n) + \alpha(g_j - f_j)$。如果 $g_j > f_j$，

则上述两个加项都是正的。如果 $g_j < f_j$，则

$$T\left(\frac{j}{n}\right) + \alpha(g_j - f_j) > T\left(\frac{j}{n}\right) + \frac{1}{n}(g_j - f_j) = T\left(\frac{j+1}{n}\right) \geqslant 0$$

由此，除了一些技术细节之外，证明结束。$\hat{F}_n$ 和 $\hat{G}_n$ 都不一定要是累积分布函数。我们对此做出补救，定义在 $x \in (0, 1)$ 区间 $F_n(x) = \hat{F}_n(x)$，且 $F_n(0) = 0$，$F_n(1) = 1$。用类似的方法定义 G_n。如果 $\hat{F}_n$ 和 $\hat{G}_n$ 满足式（11）、式（12）和式（13），则 F_n 和 G_n 也满足式（11）、式（12）和式（13）。

三、分布函数的偏序

“更大的不确定性”的定义是（或者应该是）分布函数集合的偏序的定义。在本部分，我们将正式地定义三种偏序，分别对应于我们在第一部分介绍的递增风险的前三个概念，并且证明其等价性。

1. 偏序

集合上的偏序 $\leqslant_p$ 是一个二元的，具有传递性、反身性、反对称性[8] 的关系。我们定义偏序的集合就是在定义域［0，1］上的分布函数的集合。$F \leqslant_p G$ 和 $X \leqslant_p Y$ 可以交换使用，其中 F 和 G 分别是随机变量 X 和 Y 的累积分布函数。

2. $\leqslant_I$ 的定义

延续上一部分的讨论，我们定义偏序 $\leqslant_I$ 如下：当且仅当 $G—F$ 满足积分条件（6）和条件（7）的时候，$F \leqslant_I G$。

引理 3：$\leqslant_I$ 是一个偏序。

证明：$\leqslant_I$ 具有传递性和反身性是不证自明的。我们只需要证明反对称性。假设 $F \leqslant_I G$ 且 $G \leqslant_I F$。如下定义 S_1 和 S_2：

$$S_1 = G - F \quad 并且 \quad S_2 = F - G$$

因此，$S_1 + S_2 = 0$。进而，因为 $F \leqslant_I G$ 且 $G \leqslant_I F$，如果 $T_i(y) = \int_0^y S_i(x)dx$，则 $T_i(y) \geqslant 0$。因为 $0 = \int_0^y (S_1(x) + S_2(x))dx = T_1(y) + T_2(y) = 0$ 并且 $T_i(y) \geqslant 0$，所以 $T_i(y) = 0$。我们需要证明这意味着 $S_1(x) = 0$（几乎在所有地方），或者 $F(x) = G(x)$（几乎在所有地方）。这样就能完成对该引理

8 如果 $A \leqslant_p B$ 且 $B \leqslant_p A$ 意味着 $A = B$，则称一对关系 $\leqslant_p$ 是反对称的。

的证明。[9]

由于 $S_1(x)$ 的变化是有界的（它是两个单调函数的差），其断点构成了一个测度为零的集合。我们称其为集合 N。定义

$$\hat{S}_1(x) = \begin{cases} 0 & \text{对于} \quad x \in N \\ S_1(x) & \text{其他} \end{cases}$$

由此 $\int_0^y S_1(x)dx = \int_0^y \hat{S}_1(x)dx = T_1(y)$。假设存在一个 $\hat{x}$ 满足 $\hat{S}_1(\hat{x}) \neq 0$，比如 $\hat{S}_1(\hat{x}) > 0$。则在 $x \in (\hat{x}-\epsilon, \hat{x}+\epsilon)$ 区间（对某个 $\epsilon > 0$），$\hat{S}_1(x) > 0$（因为 $\hat{S}_1(x)$ 在 $\hat{x}$ 点连续）。由此，$T_1(x-\epsilon) < T_1(x+\epsilon)$。这组矛盾完成了本引理的证明。

3. $\leqslant_u$ 的定义

我们根据如下想法定义偏序 $\leqslant_u$：如果每一个风险厌恶者相对于 Y 而言，都更偏好于 X，则称 X 比 Y 风险低。当且仅当对每一个有界凹函数 U 都有 $\int_0^1 U(x)dF(x) \geqslant \int_0^1 U(x)dG(x)$ 时，$F \leqslant_u G$。不证自明，$\leqslant_u$ 具有传递性和反身性。$\leqslant_u$ 具有反对称性是下面给出的定理 2 的直接结果。

4. $\leqslant_a$ 的定义

如果 Y 的分布和 X 加上某个噪音的分布相同，则称 X 的风险比 Y 小，与此相应的是我们此处要定义的偏序 $\leqslant_a$。$F \leqslant_a G$ 当且仅当存在随机变量 X 和 Z 一个联合分布函数 $H(x, z)$（定义在 $[0, 1] \times [-1, 1]$ 上的）满足如下条件：如果

$$J(y) = \Pr(X + Z \leqslant y)$$

则

$$F(x) = H(x,1), 0 \leqslant x \leqslant 1$$

$$G(y) = J(y), 0 \leqslant y \leqslant 1$$

并且

$$E(Z \mid X = x) = 0 \quad \text{对所有的 } x \text{ 成立} \tag{14}$$

以随机变量表示的等同定义如下：如果存在一个满足式（14）的随机变量 Z 使得

9 我们将延续惯例，如果两个分布函数仅在测度为零的集上有区别，则称两个分布函数是等同的。

$$Y \underset{d}{=} X + Z \qquad (15)$$

则称 $X \leqslant_a Y$。非常重要的一点是，读者必须了解式（15）并不意味着 $Y = X + Z$。

在 X 和 Y 是在有限点上集中的离散分布函数的时候，可以对 $\leqslant_a$ 进行非常有用并且易于处理的描述。不失一般性的情况下，我们假设 X 和 Y 在 a_1，a_2，…，a_n 点上集中。则 X 和 Y 的累积分布函数由以下数值决定：

$$f_i = \Pr(X = a_i)$$

以及

$$g_i = \Pr(Y = a_i)$$

因此 $X \leqslant_a Y$，当且仅当存在 n^2 个 $c_{ij} \geqslant 0$ 满足

$$\sum_j c_{ij} = 1, i = 1, \cdots, n \qquad (16)$$

$$\sum_j c_{ij}(a_j - a_i) = 0, i = 1, \cdots, n \qquad (14')$$

并且

$$g_j = \sum_i f_i c_{ij}, \quad j = 1, \cdots, n \qquad (15')$$

为证明这一点，我们如下定义一个条件于 X 的随机变量 Z

$$c_{ij} = \Pr(Z = a_j - a_i \mid X = a_i)$$

式（16）表明，这个方程事实上定义了一个随机变量，而式（14′）和式（15′）是和式（14）与式（15）类似的。这些条件可以以矩阵形式表示：

$$Ca = a \qquad (14'')$$

$$g = fC \qquad (15'')$$

$$Ce = e \qquad (16'')$$

其中 $e =$（1，…，1）是完全由 1 组成的向量。如果 f^1，f^2 和 f^3 是随机变量 X^1，X^2 和 X^3 的累积分布函数（$f_i^k = \Pr(X^k = a_i)$），并且如果 $X^1 \leqslant_a X^2$，$X^2 \leqslant_a X^3$，则存在矩阵 C^1 和 C^2 满足 $C^1 a = C^2 a = a$；$C^1 e = C^2 e = e$，同时 $f^2 = f^1 C^1$，$f^3 = f^2 C^2$。令 $C^* = C^1 C^2$，则 $f^3 = f^1 C^*$ 并且 $C^* a = C^1 C^2 a = C^1 a = a$，类似地，$C^* e = e$。我们已经证明了如下引理。

引理 4：如果 X^1，X^2 和 X^3 在有限点上集中，则 $X^1 \leqslant_a X^2 \leqslant_a X^3$ 意味着 $X^1 \leqslant_a X^3$。

5. $\leqslant_I$，$\leqslant_a$，$\leqslant_u$ 的等价性

我们现在给出并证明本文的主要结论。

定理2：下面的几个陈述是等价的：

(A) $F \leqslant_u G$

(B) $F \leqslant_I G$

(C) $F \leqslant_a G$

证明：证明的顺序如下：(C) ⇒ (A) ⇒ (B) ⇒ (C)。在整个证明过程中，我们沿用在第一部分结束时介绍的标记惯例。

(a) $X \leqslant_a Y \Rightarrow X \leqslant_u Y$

假设存在一个随机变量 Z 满足 $Y \underset{d}{=} X + Z$ 并且 $E(Z \mid X) = 0$。对每个固定的 X 和凹函数 U，如上对 Z 取期望，根据简森不等式：

$$E_X U(X+Z) \leqslant U(E(X+Z)) = U(X)$$

对 X 取期望，得到

$$EE_X U(X+Z) \leqslant EU(X)$$

或者

$$EU(Y) \leqslant EU(X)$$

(b) $F \leqslant_u G \Rightarrow F \leqslant_I G$[10]

如果 $S = G - F$，则 $F \leqslant_u G$ 意味着 $\int_0^1 U(x)dS(X) \leqslant 0$ 对所有凹函数 U 成立。由于密度函数及其负函数都是凹的，我们得到 $\int_0^1 xdS(x) \leqslant 0$ 和 $\int_0^1 (-x)dS(x) \leqslant 0$，因此 $\int_0^1 xdS(x) = 0$。通过部分积分我们得到 $T(1) = 0$。我们仍需证明 $T(y) \geqslant 0$ 对所有 $y \in [0, 1]$ 成立。对于给定的 y，我们令 $b_y(x) = \mathrm{Max}(y - x, 0)$。由此 $-b_y(x)$ 是凹的，并且 $0 \leqslant \int_0^1 b_y(x)dS(x) = \int_0^y (y-x)dS(x) = yS(y) - \int_0^y xdS(x)$。对最后一项进行部分积分，我们得到

$$-\int_0^y xdS(x) = -xS(x)\Big]_0^y + \int_0^y S(x)dx$$
$$= -yS(y) + T(y)$$

因此，$T(y) = \int_0^1 b_y(x)dS(x) \geqslant 0$。

(c) $F \leqslant_I G \Rightarrow F \leqslant_a G$

10　这里简便的证明形式受益于戴维·华莱士（David Wallace）的帮助。对于连续可微的函数 U，反向证明可以通过部分积分获得。

我们首先证明这一点在如下特例情况下成立：F 和 G 是相差一个均值保留展型变化的离散随机变量。使用第二部分第 2 点的标记，令 F 和 G 在 a_1，a_2，a_3，a_4（$a_1 < a_2 < a_3 < a_4$）之外的点上概率权重相同。令 $\Pr(X = a_k) = f_k$，$\Pr(Y = a_k) = g_k$。如果 $\gamma_k = g_k - f_k$，则

$$\gamma_1 = -\gamma_2 \geqslant 0, \gamma_4 = -\gamma_3 \geqslant 0 \tag{4. i}$$

以及

$$\sum_{k=1}^{4} \gamma_k a_k = 0 \tag{4. ii}$$

是 F 和 G 相差一个均值保留展型变换的条件。为证明 $F \leqslant_a G$，我们只需要证明存在 $c_{ij} \geqslant 0$（i，$j = 1$，2，3，4）满足式（14′），式（15′）和式（16）。考虑

$$\{c_{ij}\} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \dfrac{\gamma_1(a_4 - a_2)}{f_2(a_4 - a_1)} & \dfrac{g_2}{f_2} & 0 & \dfrac{\gamma_1(a_2 - a_1)}{f_2(a_4 - a_1)} \\ \dfrac{\gamma_4(a_4 - a_3)}{f_3(a_4 - a_1)} & 0 & \dfrac{g_3}{f_3} & \dfrac{\gamma_4(a_3 - a_1)}{f_3(a_4 - a_1)} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{17}$$

容易证明，式（17）定义的 c_{ij} 满足式（16）和式（14′）。因此，如果我们定义了 Z

$$c_{ij} = \Pr(Z = a_j - a_i \mid X = a_i)$$

则 Z 是一个条件于 X 的随机变量，满足 $E(Z \mid X) = 0$。我们仍需证明式（15′）或者 $Y \underset{d}{=} X + Z$。考虑 $Y^1 = X + Z$。Y^1 是一个离散随机变量，由于 $E(Z) = 0$，Y^1 和 Y 具有同样的均值。它们之间的不同仅仅在于它们可以对 a_1，a_2，a_3，a_4 点赋予不同的概率权重。但是

$$\Pr(Y^1 = a_2) = \Pr(X = a_2) \cdot \Pr(Z = 0 \mid X = a_2)$$

$$= f_2 \cdot \frac{g_2}{f_2} = g_2 = \Pr(Y = a_2)$$

类似地，$\Pr(Y^1 = a_3) = \Pr(Y = a_3)$。这样，$Y^1$ 和 Y 最多在两点上赋予不同的权重。但是 $\Pr(Y = a_1) > \Pr(Y^1 = a_1)$ 意味着 $\Pr(Y^1 = a_4) > \Pr(Y = a_4)$，而后者反过来意味着 $E(Y^1) > E(Y)$，这构建了一对矛盾。因此，$Y \underset{d}{=} Y^1 \underset{d}{=} X + Z$。

引理 1 和引理 4 使得我们可以将此结果扩展到所有的带有有限个增值点

的离散分布函数。我们将定理 1（b）扩展到所有的累积分布函数。如果 $F \leqslant_I G$，则存在序列 $\{F_n\}$ 和 $\{G_n\}$，由含有有限个增值点的离散分布函数组成，满足 $F_n \to F$，$G_n \to G$，并且 $F_n \leqslant_I G_n$。我们已经证明了 $F_n \leqslant_a G_n$。令 X_n 和 Y_n 是分布函数分别为 F_n 和 G_n 的随机变量。对于每个 n，都存在一个随机变量 X_n 和 Z_n 的联合分布函数 $H_n(x, z)$，使得如果 $J_n(y) = \Pr(X_n + Z_n \leqslant y)$，则

$$J_n(y) = G_n(y) \tag{18}$$
$$F_n(x) = H_n(x, 1) \tag{19}$$

并且

$$E(X_n \mid Z_n) = 0 \tag{20}$$

由于 H_n 是一个离散分布函数，对于所有定义于［0，1］的连续函数 u，式（20）可以表达为

$$\int_0^1 \int_{-1}^1 u(x) z dH_n(x, z) = 0 \tag{21}$$

由于 H_n 是随机有界的（stochastically bounded），序列 $\{H_n\}$ 有一个子序列 $\{H_{n'}\}$ 收敛于随机变量 X 和 Z 的分布函数 $H(x, z)$。[11] 由于 $H_{n'}(x, 1) = F_{n'}(x) \to F$，$H_{n'}(x, 1) \to F$。类似地，$J_{n'} \to G$。令

$$M_{n'} = \int_0^1 \int_{-1}^1 u(x) z dH_{n'}(x, z)$$

根据弱收敛的定义，$M_{n'} \to \int_0^1 \int_{-1}^1 u(x) z dH(x, z)$。但是 $\{M_{n'}\}$ 是一组所有的项都是 0 的序列，它必然也收敛于 0。因此 $\int_0^1 \int_{-1}^1 u(x) z dH(x, z) = 0$，这意味着 $E(Z \mid X) = 0$。由此，证明结束。

6. 进一步评论

我们将用两个关于这些排序的评论来结束本部分。

A. 偏序对全序（Complete Ordering）。在前一小节，我们已经证明了 $\geqslant_a$，$\geqslant_I$ 和 $\geqslant_u$ 对具有同样均值的分布函数，建立了等同的偏序。必须强调的是，这些排序都是偏序的，也就是说，如果 F 和 G 具有同样的均值，但是 $\int_0^1 (F(x) - G(x)) dx = T(y)$ 改变了正负号，则 F 和 G 无法排序。但是这反过来意味着总是存在着两个凹函数 U_1 和 U_2，满足 $\int_0^1 U_1 dF(x) > \int_0^1 dG(x)$

---

11 ［参考文献 3，pp. 247，261］。

同时 $\int_0^1 U_2 dF(x) < \int_0^1 U_2 dG(x)$；也就是说，存在某个风险厌恶者在 F 和 G 之间偏好 F，而另一个风险厌恶者在 F 和 G 之间偏好 G。另一方面，与均值方差分析相关的排序 $\geqslant_V$（如果 $EX = EY$，$EX^2 \leqslant EY^2$，则 $X \leqslant_V Y$）是一个全序。即如果 X 和 Y 有同样的均值，则或者 $X \leqslant_V Y$，或者 $X \geqslant_V X$。[12]

B. 凹性（Concavity）。我们已经注意到，如果 U 是凹的，$X \leqslant_I Y$ 意味着 $EU(X) \leqslant EU(Y)$。类似地，给定一个定义在［0，1］上的可微函数 U，它既不是凹的，也不是凸的，则存在分布函数 F，G，以及 H，$F \geqslant_I G \geqslant_I H$，满足 $\int_0^1 U(x) dF \leqslant \int_0^1 U(x) dG$，但是 $\int_0^1 U(x) dG \geqslant \int_0^1 U(x) dH$。

简而言之，$\geqslant_I$ 定义了所有的凹函数的集合：函数 U 是凹的，当且仅当 $X \leqslant_I Y$ 意味着 $EU(X) \leqslant EU(Y)$。

四、均值方差分析

最常用的比较不确定性期望（uncertain prospects）的方法就是均值方差分析。容易证明，这样的比较可能会得到不公正的结论。比如，如果 X 和 Y 具有相同的均值，X 的方差可能比较小，但是某些风险厌恶者可能在 X 和 Y 之间更偏好 Y。为证明这一点，我们只需要观察到，尽管 $F \leqslant_u G \Rightarrow F \geqslant_V G$（因为方差是一个凸函数），但 $F \geqslant_V G$ 并不意味着 $F \geqslant_u G$。事实上，用与前文类似的证明，可以证明 U 是二次函数，当且仅当 $X \geqslant_V Y$ 意味着 $EU(X) \geqslant EU(Y)$。由此得到的一个直接结论是：如果 $U(x)$ 是一个非二次的凹函数，则存在具有相同均值的随机变量 X_i（$i = 1, 2, 3$），满足 $EX_1{}^2 < EX_2{}^2$，但是 $EX_2{}^2 > EX_3{}^2$，同时 $EU(X_1) < EU(X_2) < EU(X_3)$。即以方差排序和以期望效用排序是不同的。

托宾曾经推测，如果分布的种类——从而分布的变化种类——受到限

12 要得到这一结论，还有一种方法就是观察到 $\geqslant_V$ 比 $\geqslant_I$ 更强，因为很多能够根据 $\geqslant_V$ 排序的分布函数不能由 $\geqslant_I$ 排序。很明显，也存在着比 $\geqslant_I$ 更弱的排序。在我们的以前版本的论文中，曾经关注过一种更弱的排序：随机变量 X 是随机变量 Y 和一个确定量——一个集中在 $E(Y)$ 点的随机变量——的混合，它和确定量有着同样的均值，则 X 一定是比 Y 风险性小的。我们可以由此定义一个偏序 $\geqslant_M$。很明显，由 $\geqslant_M$ 可以推到 $\geqslant_I$，因为 X 和 Y 之间的差满足积分条件。同样清楚的是，$\geqslant_M$ 是一个很弱的偏序，因为只有很少的随机变量可以用 $\geqslant_M$ 排序。事实上，如果 $\overline{Y}$ 是集中于 $E(Y)$ 的确定值，则可以证明 $Y \geqslant_M X$ 当且仅当在 $0 \leqslant a \leqslant 1$ 上，$X \underset{d}{=} aY + (1-a)\overline{Y}$。这表明 $\geqslant_M$ 并不是一个非常让人感兴趣的偏序。我们需要感谢一位匿名评审者指出 $\geqslant_M$ 的不足之处。

制，均值方差分析就是合适的。这是正确的，但是据目前所知，这个限制是非常严格的。托宾的证明——正如他暗中意识到的（[参考文献7, pp. 20－21]）——仅仅对只相差“位置参数”（location parameters）的分布是有效的（对这一经典概念的讨论可以参见［参考文献3, p. 144］）。也就是说，托宾只考虑了这样一种分布的变化——分布从 F 变化为 G，存在 a 和 b $(a>0)$ 使得 $F(x)=G(ax+b)$。这样的变化只是分布的中心的变化，以及分布的统一收缩或扩张——和单位的变化等同。

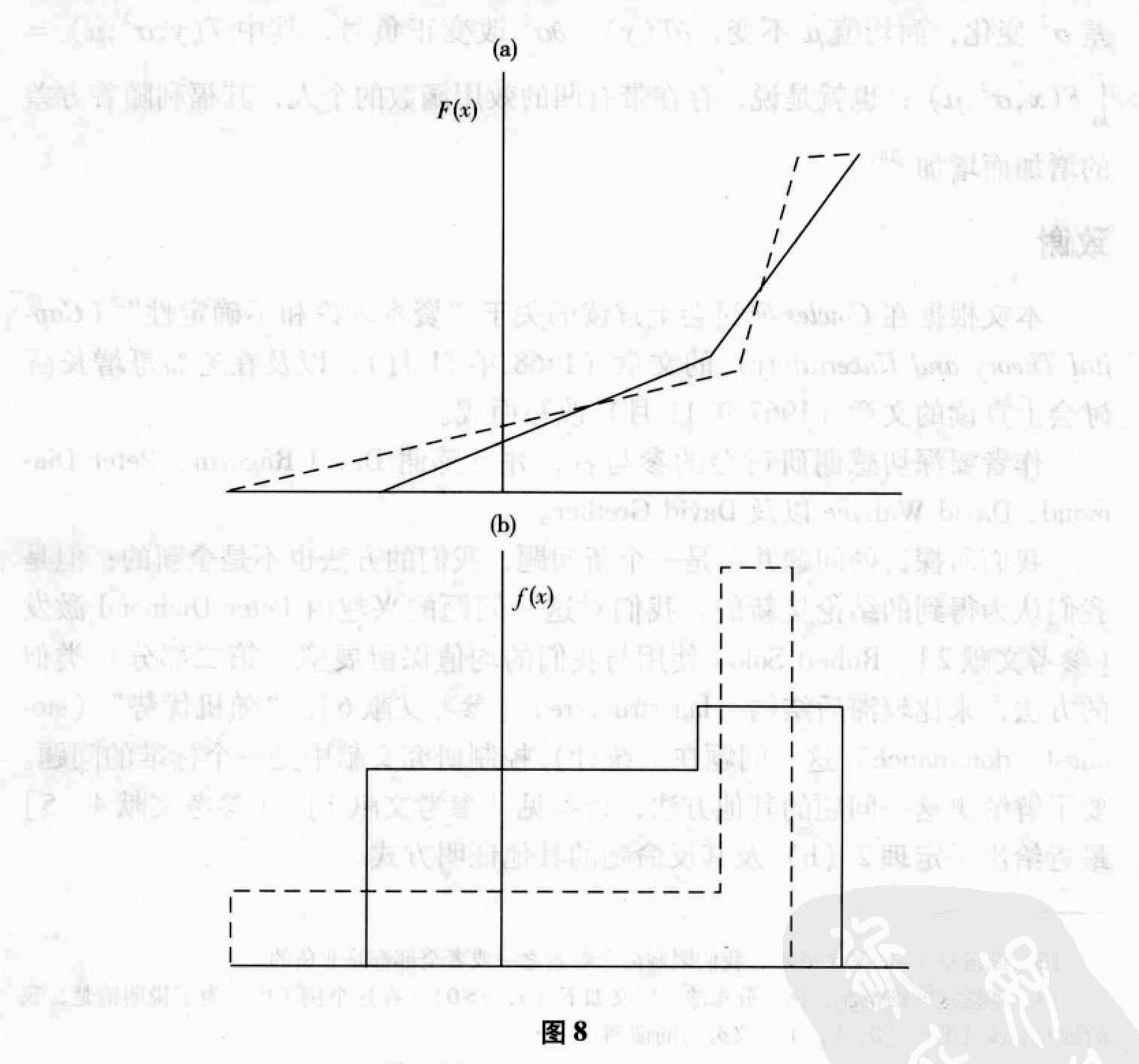


图8

关于分布函数的双参数族（parameter family）的概念，存在一些没有必要的混淆。不可否认的是，所有仅相差位置参数的分布构成了一个双参数族。大体而言，“双参数族”指的是什么？对我们而言，分布的双参数族是这样一种分布的集合：该集合中的一个分布将被挑选出来以确定两个参数。

正如托宾所指出的，双参数族是这样一种分布的集合："只要知道两个参数就可以描述整个分布。"从技术角度来说，双参数族就是从 E^2 到分布函数空间的映射。[13]非常清楚，对于这个双参数族的宽泛定义，托宾的推测不可能成立，因为对此映射的范围没有任何限制。

对双参数族采用其他定义也是可能的。它们本质上都涉及对从 E^2 到分布函数空间的"良好"映射的限制。例如，一族带有外在代数形式的，仅包含两个可以变化的参数的分布。然而，我们很容易构建如下这个例子：方差 σ^2 变化，而均值 μ 不变，$\partial T(y)\ /\partial\sigma^2$ 改变正负号，其中 $T(y,\sigma^2,\mu)=\int_0^y F(x,\sigma^2,\mu)$；也就是说，存在带有凹的效用函数的个人，其福利随着方差的增加而增加。[14]

致谢

本文根据在 *Cowles* 研讨会上宣读的关于"资本理论和不确定性"（*Capital Theory and Uncertainty*）的文章（1968 年 11 月），以及在芝加哥增长研讨会上宣读的文章（1967 年 11 月）改写而成。

作者要深切感谢研讨会的参与者，并且感谢 David Ragozin，Peter Diamond，David Wallace 以及 David Grether。

我们所探讨的问题并不是一个新问题，我们的方法也不是全新的；但是我们认为得到的结论是新的。我们对这一问题的兴趣由 Peter Diamond 激发［参考文献 2］。Robert Solow 使用与我们的均值保留展型（第二部分）类似的方法，来比较滞后结构（lag structure）［参考文献 6］。"随机优势"（stochastic dominance）这一问题在（统计）控制研究文献中是一个标准的问题。要了解解决这一问题的其他方法，请参见［参考文献 1］。［参考文献 4、5］最近给出了定理 2（b）及其反命题的其他证明方式。

---

13 或者是 E^2 的某个子集；我们限制两个参数之一或者全部都是非负的。

14 考虑这样的情况，比如分布的族定义如下（$a,\ c>0$）（在这个例子中，为了说明清楚，我们放弃了通常的在［0，1］上定义分布的惯例）。

$$F(x;a,c)=\begin{cases}0 & \text{对于 } x\leqslant 1-0.25/a\\ ax+0.25-a & \text{对于 } 1-0.25/a\leqslant x\leqslant 1+(2c-0.5)/(c-a)\\ cx+0.75-3c & \text{对于 } 1+(2c-0.5)/(c-a)\leqslant x\leqslant 3+0.25/c\\ 1 & \text{对于 } x>3+0.25/c\end{cases}$$

这个族中的两个分布具有相同的均值，但是不同的方差，如图 8（a）所示。它们显然不满足条件（7）。它们的密度函数如图 8（b）所示。

参考文献

1. S. A. BESSLER AND A. F. VEINOTT, JR., Optimal policy for a dynamic multi-echelon inventory model, *Naval Research Logistics Quarterly* **13** (1966), 355–387.
2. P. A. DIAMOND, Savings decisions under uncertainty, Working paper no. 71, Institute of Business Economic Research, University of California, Berkeley, June 1965.
3. W. FELLER, "An Introduction to Probability Theory and Its Applications," Vol. II, Wiley, New York, 1966.
4. J. HADER AND W. RUSSELL, Rules for ordering uncertain prospects, *Amer. Econ. Rev.* **59** (1969), 25–34.
5. G. HANOCH AND C. LEVY, Efficiency Analysis of Choices Involving risk, *Review of Economic Studies* **36** (1969).
6. R. M. SOLOW, A note on dynamic multipliers, *Econometrica* **19** (1951), 306–316.
7. J. TOBIN, The theory of portfolio selection, *in* F. Hahn and F. Brechlıng, "The Theory of Interest Rates," MacMillan, London, 1965.
8. J. TOBIN, Comment on Borch and Feldstein, *Rev. Econ. Studies* **36** (1969), 13–14.

递增的风险：经济意义 *

一、绪论

在前一篇文章［参考文献 11］，我们证明了如下三个命题的等价性。这些命题都是在对具有相同期望价值的随机变量 X 和 Y 进行比较：

（a）“所有的风险厌恶者——那些具有凹的效用函数的人——在 X 和 Y 之间都更偏好 X。”

（b）“Y 等于 X 加上某个噪音。”

（c）“Y 在尾部的权重比 X 在尾部的权重大。”[1]

对我们而言，这三个明显不同的概念的等价性很好地证明了，我们对递增风险的定义是一个自然的定义。它同时给出了研究风险对于经济决策的效应的多种方法。本文研究两种这样的方法。

第一种方法关注递增风险的经济效应。本文包含了一些对用传统均值方差方法解决这一问题的尖锐批评。[2] 请注意，要使某个建议是完美的，就需要证明它有可能臻于完美，我们将致力于展示我们的方法可以产生“经济学家们感兴趣的比较静态结果［参考文献 16］”。本文的下一部分将努力实

---

* “Increasing Risk II：Its Economic Consequences，” with M. Rothschild，Journal of Economic Theory，5（1），March 1971，pp. 66-84。本文的研究课题得到了美国国家科学基金会和福特基金会的资助。

1 （a），（b）和（c）的正式表述是：

（a）对于所有凹的 Y，$EU(X) \geqslant EU(Y)$。

（b）存在随机变量 Z 使得 Y 和 $X+Z$ 具有同样的分布，其中对所有的 X，都有 $E[Z \mid X]=0$。

（c）如果 F 和 G（分别是 X 和 Y 的分布函数）的增值点，都限在闭区间［a，b］上，那么如果 $T(y) = \int_a^Y (F(x) - G(x))dx$，则 $T(y) \geqslant 0$ 并且 $T(b) = 0$。

2 正如我们在“递增的风险：一个定义”中强调的那样，我们的方法仅仅产生了分布的偏序。均值方差方法产生了分布（具有相同的期望价值）的全序，因此，通常能对不确定性改变产生的效应给出解答，但我们的方法却不可以。但是这一点是对我们的方法的支持，而不是驳斥。均值方差分析给出的解答是虚假的，当且仅当效用函数或者分布的种类受到严格限制的时候，它们才成立。另外，均值方差分析对于必须给出怎样的限制才能使其结论成立并没有线索。正如我们将在后文中再次证明的那样，必要限制条件是我们的分析方法的副产品。

现这一承诺。

如果个人的效用函数带有某个控制参数 α，以及一个随机变量 θ，个人将通过选择 α 来最大化其期望效用

$$\max_{\alpha}\int U(\theta,\alpha)\,dF\theta \tag{1}$$

最优的 α 一定满足下列条件：

$$\int \frac{\partial U(\theta,\alpha)\,dF(\theta)}{\partial\alpha} = EU_{\alpha}(\theta,\alpha) = 0 \tag{2}$$

令 α^* 是式（2）的唯一解，并且假设在 α^* 的邻域，U 对 α 是单调递减的。如果 $U_{\alpha}(\theta,\ \alpha)$ 是对 θ 的凹函数，风险性的增加将导致 α^* 值的减小。之所以出现这样的结果，是因为（a）严格表明风险性的增加将减少 $EU_{\alpha}(\theta,\ \alpha)$；降低 α 重新构造了式（2）的一阶条件。类似地，如果 $U_{\alpha}(\theta,\ \alpha)$是对 θ 的凸函数，θ 易变性的增加会使 α^* 的值增大。如果 $U_{\alpha}(\theta,\ \alpha)$既不是凸的，也不是凹的，则风险性增加引起的效应就是模糊的。[3]

本文的下一部分将应用这一想法。我们试图确定在什么样的条件下，在简单的经济模型中产生的函数是凹的或是凸的。前三个模型关注储蓄和投资行为，而后三个模型关注在生产或者需求条件不确定的情况下的技术选择和投入水平选择问题。[4] 我们的主要结论是：均值方差方法得出了误导性的一般结论，并且用阿罗—普拉特（Arrow - Pratt）［参考文献 1、9］的绝对和相对风险厌恶系数来表述相关函数为凸或凹的条件是很有益的。当这一方法不能对风险增加的作用给出确定的答案时，很自然的问题是：风险的什么样的增加可以有确定性的作用。我们发现强调效用函数和分布函数变化之间互动关系（这正是前一篇文章中积分条件的基础）通常可以给出解决这一问题的有用方法。相关的一个例子就是下文的投资组合问题。

---

3 早期的研究［参考文献 4、14 等］曾经对完全确定性和不确定性情况做过比较。在“确定”的情况下，我们选择 α 来使得

$$U_{\alpha}(X,\hat{\alpha}) = 0$$

$\hat{\alpha}\gtreqless\alpha^*$（$\alpha^*$ 是式（1）的解）取决于 $EU_{\alpha}(X,\ \hat{\alpha})\gtreqless U\ (EX,\ \hat{\alpha})$，当 U_{α} 对 X 是凸的或是凹的时候，简森不等式都允许我们给出清楚的陈述。同样在这一条件下，我们可以对更多种类的问题做出明确的陈述。

4 在［参考文献 11］中，这些概念的进一步重要应用，是关于收入不平等问题的，这个问题在 A. B. Atkinson［参考文献 2］中已有详细讨论。

本文对［参考文献 11］的结果的有用性做了另一不同的证明，并以此作为本文的结论。不确定性下的选择可以看作是随机变量的选择。例如，投资组合的选择可以看作是随机回报流的选择。为了与上文讨论的观点对比，我们将个人的问题构建为选择 β 以最大化

$$\int U(x)dF(x,\beta) \tag{3}$$

即，他要选择最有利的随机变量。条件（b）给出了一个标准，以此我们可以决定风险厌恶者偏好于哪个随机变量。本文的最后一个部分给出了两个例子，从中可以看出这个条件如何被使用。

二、递增风险的作用

1. 储蓄和不确定性

我们从分析储蓄回报率的不确定性有什么作用，开始我们的讨论。[5] 假设个人拥有给定的财富 W_0，他要将 W_0 分配于今天的消费和明天的消费。不在今天消费掉的那部分财富被用于投资；在期末，投资的每一单位美元收益是随机回报 e。他要将 W_0 在两期之间分配，以最大化两期的期望效用：

$$E[U(C_1)+(1-\delta)U(C_2)] = U((1-s)W_0)+(1-\delta)EU(sW_0e) \tag{4}$$

其中 s 是储蓄率，δ 是净的时间折现率（pure rate of time discount）。假设个人是风险厌恶的，则效用函数满足下列条件：

$$U'>0;\quad U''<0 \tag{5}$$

效用最大化的必要条件（因为式（5），这些条件也是充分条件）是式（4）相对于 s 的导数为零。最优储蓄率一定满足：

$$U'((1-s)W_0) = E[U'(sW_0e)](1-\delta)e \tag{6}$$

直觉（以其一贯的模棱两可）告诉我们，储蓄回报的不确定性的增加可能减少储蓄（因为“一鸟在手，胜过双鸟在林”），也可能增加储蓄（因为风险厌恶的个人为了保证其最低生活标准会在不确定性增加时储蓄更多）。均值—方差分析——等同于假设效用函数是二次型的——表明，只有第一个论述是正确的。如果 $U(C) = aC-\frac{1}{2}bC^2$，那么式（6）的右边等于：

$$(1-\delta)(aE(e)-bsW_0E(e^2))$$

5　更详细的讨论以及一些相关的问题［参考文献 4、5、7、14］。

随着 e 变得更具风险，此项会递减。s 的减小可以弥补这对一阶条件的扰动——s 的减小可以使式（6）的左边减小，使式（6）的右边增大。

但这一结果并不一般化。递增的易变性使得 s^* 增大还是减小，这取决于 $eU'(sW_0e)$ 是 e 的凹函数还是凸函数。因此，如果对于所有的正数 C：

$$2U''(C) + U'''(C)C > 0 \tag{7}$$

则递增的风险可以增加储蓄；如果对于所有的正数 C：

$$2U''(C) + U'''(C)C < 0 \tag{8}$$

则递增的风险可以减少储蓄。否则，风险增加的作用是模棱两可的。显然，递增的风险使得储蓄减少的一个充分条件是三阶导非正。阿罗—普拉特的相对风险厌恶系数（$R = -U''C/U'$）可以用来以稍微不同的形式表述这些结果。因为 R' 的符号与 $-(U'''C + U''(1+R))$ 的符号一致，如果 R 是非增的并且大于1，那么式（7）成立；如果 R 是非减的并且小于1，那么式（8）成立。因此对于具有不变相对风险厌恶系数的一类效用函数，$U(W) = (1-a)W^{1-a}$（$a>0$，$a\neq 1$），升高了的风险是增加还是减少了储蓄取决于 $a>1$ 还是 $a<1$。[6]

2. 投资组合问题

传统的投资组合问题采用了类似的（尽管也是更复杂的）分析。假设个人需要将其投资分配于两种资产：货币（收益为零）和一种风险资产（其随机回报率为 e）。如果他将其初始财富 W_0 中的 α 投资于风险资产，那么他的最终财富为 $W(\alpha) = W_0(\alpha e + 1)$。最优的投资组合要使得最终财富的期望效用 $EU(W(\alpha))$ 最大，[7] 其中 U 是满足式（5）的递增凹函数。因此，所选定的 α 是下式的一个解：

$$H(\alpha) = W_0 EU'e = W_0\int U'(W(\alpha))e\,dF(e) = 0 \tag{9}$$

此处的 F 是 e 的分布函数。请注意，式（9）是一个充分必要条件，因为式（5）意味着

6 阿罗［参考文献 11］论证过，相对风险厌恶系数一般是递增的，并且——对于低收入者——其值小于1。Stiglitz［参考文献 13］对此表示了疑问。

7 因为在这一点下面，我们总是假设 W_0 是常数，因此我们是用回报率（$r = (W - W_0)/W_0$）（就像托宾所采用的）还是用最终财富的价值来构建这一问题都没有关系。如果效用被看作是 r 的函数，而不是 W 的函数，那么效用函数一些特征（如弹性）的解释会有差别（在前一例中，因为 r 可以是正的，也可以为负的，在此 U' 不可能有不变弹性）。

$$H'(\alpha) < 0 \tag{10}$$

当 e 变得更具风险时，对于风险资产的需求（以 α 表示）会如何变化？均值—方差分析再一次给出了错误的一般结论——风险资产易变性的增加使得风险资产吸引力下降，导致需求下降；这正好与常识的判断一致。如果 U 的函数形式是 $U(W) = aW - \frac{1}{2}bW^2$，那么式（9）可写为：

$$0 = W_0[(a - bW_0)E(e) - \alpha bW_0E(e^2)]$$

因此，$\alpha = (a - bW_0)E(e)/E(e^2)bW_0$。当 e 变得更具风险时，$E(e^2)$ 递增，$E(e)$ 不变；α 必定上升。

一般情况下，这是不正确的。如果 e 的分布函数从 F 变到了更具易变性的 G，那么新的投资配置系数 $\tilde{\alpha}$ 满足 $\int U'(W(\tilde{\alpha}))edG(e) = 0$。因此，当 $\int U'(W(\alpha))edSe \gtreqless 0$ 时，$\tilde{\alpha} \gtreqless \alpha$，其中 $S = G - F$。如果递增的风险会减少所有风险厌恶投资者对风险资产的需求，那么，当 $V(e) = U'(W(\alpha))$ 且 F 和 G 的增值点（points of increase）都在 (a, b) 中时，我们必定有：

$$\int_a^b V(e)edS(e) \leqslant 0 \tag{11}$$

对于所有正的、递减的 V 以及所有满足前一篇文章中积分条件的 S 都成立；即所有 S 满足：

$$\int_a^b S(e)de = \int_a^b edS(e) = 0 \tag{12a}$$

以及

$$\int_a^t S(e)de \geqslant 0, \quad 对于所有 \quad t \in (a,b) \tag{12b}$$

利用式（12b）以及微积分的第二中值定理，我们发现存在 $c \in (a, b)$，使得式（11）中的积分等于 $[V(a) - V(b)]h(c)$，其中 $h(c) = \int_a^c edS(e)$。因此，式（11）的一个充分条件是：对于所有 $c \in (a, b)$

$$h(c) \leqslant 0 \tag{13}$$

这同时也是必要条件。假设存在 $\bar{c}$ 使得 $h(\bar{c}) > 0$。为满足式（11），我们必须有 $\int_a^b V(e)dS(e) < 0$ 对于所有正的、递减的 V 都成立。一个这样的 V 可以是：

$$\bar{V}(e) = \begin{cases} V_1 & 对于 \quad a \leqslant e < \bar{c} \\ V_2 & 对于 \quad \bar{c} \leqslant e \leqslant b \end{cases}$$

其中 $V_1 > V_2 > 0$。因而，$\int_a^b \bar{V}(e)dS(e) = (V_1 - V_2)h(\bar{c}) > 0$。

这里我们仍然假定了递增的风险减少了对风险资产的需求。我们有可能证明对于递增的、凹的效用函数这是正确的，同时也可以证明没有这样的效用函数可以使得递增的风险总是使 α 增加。如果 $Z(e) = eU'(W(\alpha))$ 是 e 的凹函数，那么风险的增加总是降低对风险资产的需求。因为

$$Z''(e) = [(1 - R + W_0A)U'' + (W_0A' - R')U']W_0a$$

其中 $R = -U''W/U'$，$A = -U''/U'$（绝对风险厌恶系数）；不确定性的上升可以增加无风险投资的一个充分条件是：相对风险厌恶系数不大于 1 且非减，绝对风险厌恶系数是非增的。贝努里（Bernoulli）或对数效用函数，以及其他所有具有不变相对风险厌恶系数且系数小于 1 的效用函数都满足这些条件；其他具有不变相对风险厌恶系数的效用函数不满足这些条件。

在风险资产的易变性升高时，风险厌恶的投资者不会总是增加持有的风险资产。为证明这一论断，我们只要给出一种满足式（13）的风险增加。其中之一是：

$$\hat{S}(e) = \begin{cases} S_1 > 0 & \text{对于} \quad e_1 \leq e < 0 \\ S_2 > 0 & \text{对于} \quad 0 \leq e \leq e_2 \\ 0 & \text{其他} \end{cases}$$

其中选择的 e_i 和 S_i 使得式（12b）得到满足。此类推理过程可以用来证明赌博（这一重要的资产）中的任何一类的风险增加总是使得需求减少。

赌博也是一种投资，它只有两种结果。如果一切顺利，每一美元赌注可以获得 $1 + E(e) + \gamma$。否则只能收回 $1 + E(e) - \lambda$。我们令赌博的收益为 γ，令赌注的损失为 λ，$E(e)$ 为期望回报率。一个“公平的”赌博满足 $E(e) = 0$。风险厌恶者不会对公平赌注赌博感兴趣，而 $E(e) - \lambda \geq 0$ 的赌注表示根本不存在风险，因此，令人感兴趣的投资组合问题是：

$$E(e) - \lambda < 0 < E(e) \tag{14}$$

如果 P 是赢得赌博的概率，那么 $P\gamma - (1-P)\lambda = 0$，或者 $\lambda = [P/(1-P)]\gamma$。具有相同回报的赌博可以由（$P$，$\gamma$）空间的点来表示（图 1）。

图 1 以图表的形式给出了用我们的标准和用方差的标准来排列风险的差别。B 点表示一个可能收益为 γ_B（获得此收益的概率为 P_B）的赌博。直接应用均值保留展型（mean preserving spread），或者是前一篇文章中的积分条件，我们可以证明赌博 C（收益 γ_C 为，损失是 λ_C）比 B 风险更高，当且仅当：

$$\gamma_C > \gamma_B \tag{15a}$$

和

$$\lambda_C > \lambda_B \tag{15b}$$

显然，在 B 点右边的所有赌博都满足式（15a）。图 1 中的实线表示所有损失为 λ_B 的赌博。此线的斜率是 $d\gamma/dP\left|_{\lambda=\lambda_B}\right. = -\gamma/P\ (1-P)$。只有在此线之上的赌博才满足式（15b）。因此，只有在区域Ⅲ和Ⅳ的点才比 B 更具风险性；类似地，区域Ⅰ和Ⅵ中的点风险更小。根据我们的标准，在区域Ⅱ和Ⅴ中的点无法跟 B 作比较。

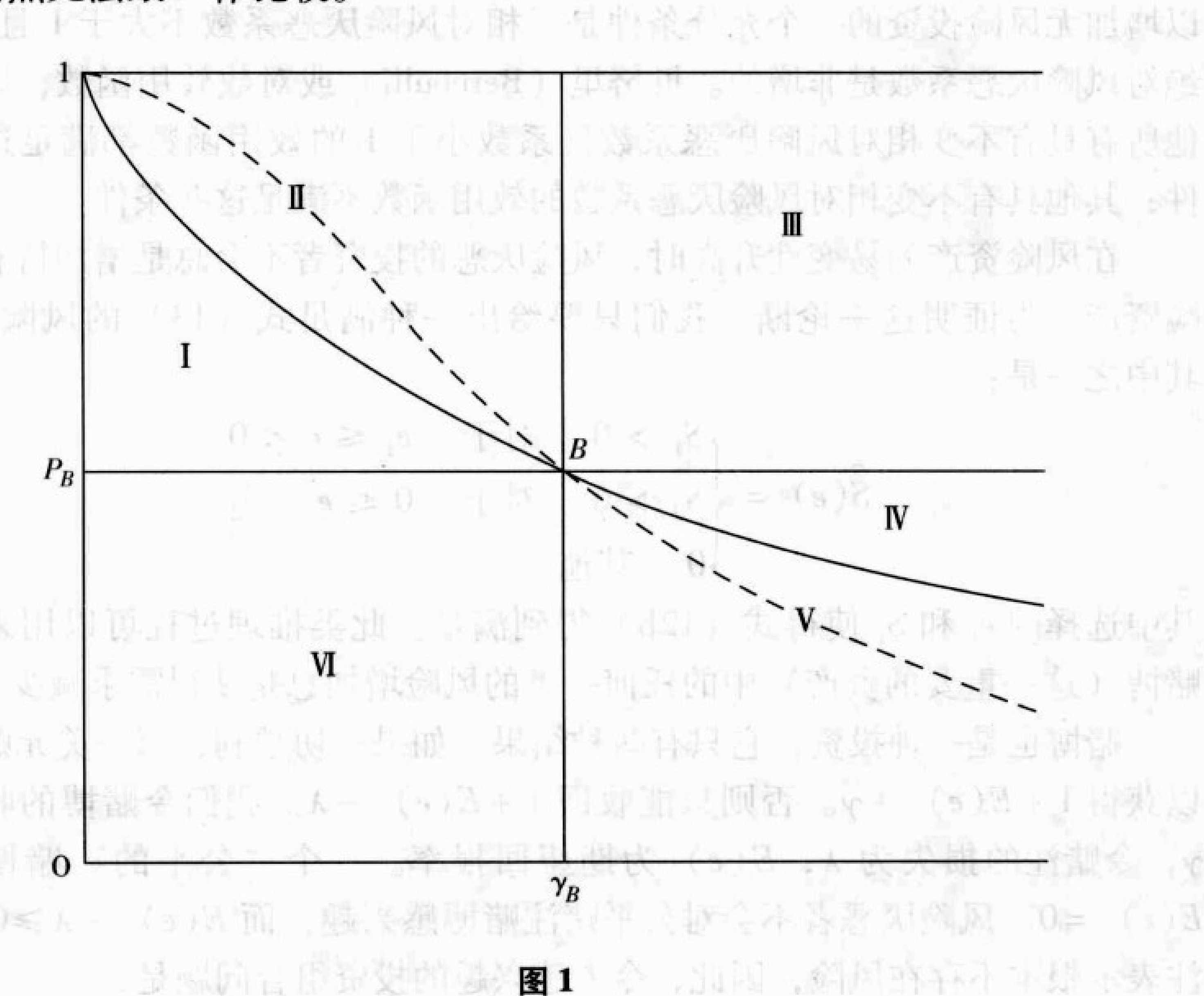


图 1

方差标准可以将图 1 中的所有点进行排序。一个赌博的方差是 $\sigma^2 = P\gamma^2 + (1-P)\ \lambda^2 = \gamma^2\ (P/\ (1-P))$。图中的虚线是方差跟 B 相同的赌博。在虚线之上的所有赌博方差都大于 B；在虚线之下的所有赌博方差都小于 B。它的斜率是：

$$\left.\frac{d\gamma}{dP}\right|_{\sigma^2=\sigma_B^2} = -\frac{1}{2}\frac{\gamma}{P(1-P)} = \frac{1}{2}\left.\frac{d\gamma}{dP}\right|_{\lambda=\lambda_B}$$

虚线从下部相交断线于 B 点，并且都在区域Ⅱ和Ⅴ中。在这两个区域中，都有一些赌博的方差大于 B，但并不比 B 更具风险。

直接利用式（13），我们可以证明，如果 C 比 B 更具风险，风险厌恶者总是愿意在其投资组合中持有更多的 B，较少的 C。[8] 引入 e_i，$i=1$，2，3，4，定义如下：

$$e_1 = E(e) - \lambda_C \qquad e_2 = E(e) - \lambda_B$$
$$e_3 = E(e) + \gamma_B \qquad e_4 = E(e) + \gamma_C$$

合并式（14）和式（15），我们可得到

$$e_1 < e_2 < 0 < e_3 < e_4 \tag{16}$$

因为 B 和 C 的期望回报相同，所以

$$(1-P_C)e_1 - (1-P_B)e_2 - P_B e_3 + P_C e_4 = \sum_{i=1}^{4} S_i e_i = 0 \tag{17}$$

在这一情形下，式（13）就变成了

$$h_k = \sum_{i=1}^{k} S_i e_i \leqslant 0; \quad k = 1,2,3,4 \tag{18}$$

因为 $e_1<0$，$S_1 = (1-P_C) >0$ 和 $h_1<0$，式（17）表明 $h_4=0$，并且 $h_3 = h_4 - P_C e_4 = -P_C e_4 <0$。我们只需进一步证明 $h_2 \leqslant 0$，或者

$$(1-P_C)e_1 - (1-P_B)e_2 = -(P_C e_4 - P_B e_3) \leqslant 0 \tag{19}$$

但是，$P_B \leqslant P_C$ 和 $P_B > P_C$ 仅有一个成立。如果 $P_B \leqslant P_C$，那么式（16）意味着式（19）的右边为负。如果 $P_B > P_C$，那么 $(1-P_B) < (1-P_C)$ ——结合式（16），这意味着式（19）的左边为负。

3. 投资组合与储蓄相结合的问题

乐弗哈里和斯利尼瓦森（Levhari and Srinivasan）[9]［参考文献5］最近分析了个人的最优投资组合和储蓄问题，其中的个人最大化消费带来的期望效用折现值，即

$$E\sum_{t=0}^{\infty}(1-\delta)^t U(C_t) \tag{20}$$

---

8 当然，这只是在 B 和一个无风险资产的一个组合与 C 和一个无风险资产的一个组合之间的比较。

满足式（13）的那些变化由前一篇文章的第一部分提到的偏序（partial ordering）$\leqslant_m$ 给出。如果 Y 是均值为 $\bar{Y}$ 的随机变量，X 是与 $\alpha Y+(1-\alpha)\bar{Y}$ 具有相同分布的随机变量，$0\leqslant\alpha\leqslant1$，那么我们记 $X\leqslant_m Y$。可以很直接地证明，如果 $X\leqslant_m Y$，那么两者累积分布函数之间的差别满足式（13）。我们之所以详细讨论了赌博的情形，原因之一在于其重要性，原因之二在于递增风险的不同概念以图示给出（在图1中给出）有其内在的好处。

9 以下简称 LS。关于动态投资组合问题的其他讨论，请参见［参考文献4］。

其中，δ是折现率，C_t是t时的消费，受限于下列随机约束：

$$W_{t+1} = (W_t - C_t) r^t ; \quad W_t \geqslant C_t \geqslant 0 \tag{21}$$

W_t是个人在t时的财富，$r^t - 1$是t时期的投资（$W_t - C_t$）的随机回报率。在t时，个人可以投资于资产1（回报率为$r_1^t - 1$），也可以投资于资产2（回报率是$r_2^t - 1$）。记α为投资于第一种资产的比例，$(1-\alpha)$是投资于第二种资产的比例，那么r^t为：

$$r^t = \alpha {r_1}^t + (1-\alpha) {r_2}^t \tag{22}$$

显然，如果：(a) r_j^t随时间为独立同分布，[10] (b) 效用函数为不变相对风险厌恶系数形式：

$$U(C) = \frac{C^{1-a}}{1-a} \quad \begin{matrix} a > 0 \\ a \neq 1 \end{matrix} \quad \text{或} \quad U(C) = \ln C \tag{23}$$

储蓄决策，C_t的决定（给定W_t）独立于资产组合的选择（α的选择）。并且，储蓄率（C_t / W_t）是独立于W_t的常数，即最优的行为采用了下述储蓄法则：将固定比例的财富储蓄起来，对于所有$W_t > 0$，

$$C_t = (1-s) W_t \tag{24}$$

LS证明，对于任何选定的α，最优的储蓄率$s(\alpha)$[11]一定满足：

$$s(\alpha)^a = (1-\delta) E(r^{1-a}) \tag{25}$$

其中

$$r = \alpha r_1 + (1-\alpha) r_2 \tag{26}$$

他们同样证明了，对于任何初始财富W_0，当选定了s使得式（25）得到满足时，最优储蓄得到的折现期望效用为：

$$V(W_0, \alpha) = \frac{1}{1-a} E \sum (1-\delta)^t C_t^{1-a} = \frac{(1-s(\alpha))^{-a} W_0^{1-a}}{1-a} \tag{27}$$

从式（25）和式（26）可以得出，s是α的函数，同样式（27）的右边也是α的函数。因此，个人通过选择α来最大化式（27）。为得到内点解，$0 < \alpha < 1$，我们必须有：

$$0 = \frac{\partial V}{\partial \alpha} = \frac{\partial V}{\partial s} \frac{\partial s}{\partial \alpha}$$

其中

---

10 这一假设允许我们在下列多数讨论中省略r_j^t的上标。

11 我们假设$a \neq 1$。$a=1$的情形需要特殊的讨论，这在下文中详细讨论。LS看来没有注意到他们的分析在$a=1$时并不成立。

$$\frac{\partial V}{\partial s} = \frac{a(1-s)^{-(1+a)} W_0^{1-a}}{1-a} \gtrless 0 \quad 其中 a \lessgtr 1$$

我们可以证明，在 α^*（使得 $\partial s/\partial\alpha = 0$）的邻域内，对于 $a \neq 1$，$(1-a)\partial^2 s/\partial\alpha^2 < 0$；因此，式（27）驻点（critical point）一定是相对最大点，这意味着只有唯一的驻点。[12]因此，条件 $\partial s/\partial\alpha = 0$ 决定了 α^*。

一个很自然的问题是，其中一个资产回报的易变性增加对于投资组合的选择和储蓄有什么影响？*LS* 证明了，在特殊的情形下，如果其中一个资产的方差增加了，而均值保持不变，那么投资于该资产的比例会下降。为考察这在一般情况下是否成立，我们注意到，对于 $a \neq 1$，$\partial s/\partial\alpha = 0$ 当且仅当

$$E(r_1 - r_2)(\alpha r_1 + (1-\alpha) r_2)^{-a} = 0 \tag{28}$$

当且仅当 $H(r_1) = (r_1 - r_2)(\alpha r_1 + (1-\alpha) r_2)^{-a}$ 是 r_1 的凸函数或者凹函数时，易变性增加的作用不是模棱两可的。但是

$$H''(r_1) = -\{\alpha(1-a) r_1 + r_2(\alpha(1+a) + 2(1-\alpha))\}$$
$$\times [a\alpha(\alpha r_1 + (1-\alpha) r_2)^{-(a+2)}]$$

当 $a < 1$ 时（假设 $\alpha > 0$），这一表达式为负，因此，r_1 易变性的增加减少了对该风险资产的需求。当 $a = 1$ 时，这一结果同样成立，但是论证会更复杂一些。对于 $a > 1$，$H''(r_1)$ 的符号是不确定的；易变性的增加可能有相反的作用。[13]

相比分析易变性变化对资产组合配置的影响，分析易变性变化对储蓄率的影响更加容易一些。从式（25）可看出，s 的变化取决于下式的变化：

$$E\{\alpha r_1 + (1-\alpha) r_2\}^{1-a} \tag{29}$$

此处 α 为最优之选择。但是，因为从式（28）可得：

$$\frac{\partial E(\alpha r_1 + (1-\alpha) r_2)^{1-a}}{\partial\alpha} = (1-a) E(r_1 - r_2)(\alpha r_1 - (1-\alpha) r_2)^{-a} = 0$$

我们仅需要考虑，在给定 α 的值时，式（29）会发生什么变化。但是

$$\frac{\partial^2(\alpha r_1 + (1-\alpha) r_2)^{1-a}}{\partial(r_1)^2}$$
$$= \alpha^2 (1-a) a [\alpha r_1 + (1-\alpha) r_2]^{-(1+a)} \gtrless 0 \quad 其中 a \lessgtr 1$$

从中可以立即得出，r 的易变性增加使得储蓄率在 $a < 1$ 时增加，在 $a > 1$ 时减少。

---

12 *LS* 对此的论证是错误的，因为他们对 $\partial^2 S/\partial\alpha^2$ 的衡量出错了。

13 这使得［参考文献 5］中从对数正态分布得出的结果一般化了。

对数效用函数（$a=1$）的情形需要我们特殊关注。在此情形下，我们可以将公式写为：

$$V(W_0) = \ln(1-s)W_0 + (1-\delta)EV(sW_0 r) \tag{30}$$

式（30）的解为：[14]

$$V(W_0) = \frac{1}{\delta}\ln(1-s)W_0 + \frac{1-\delta}{\delta^2}[\ln(s) + E\ln r] \tag{31}$$

s 的最优值由下式给出，并且独立于 r：[15]

$$s = 1-\delta \tag{32}$$

要算出最优的投资组合配置（α 的最优值），我们通过选择 α 来最大化 $V(W_0)$；即从式（31）出发，我们必须最大化 $E\ln r$。这样，α 满足：

$$E\left\{\frac{r_1 - r_2}{\alpha r_1 + (1-\alpha)r_2}\right\} = 0 \tag{33}$$

因为

$$\frac{d^2 E\ln r}{d\alpha^2} = -E\left\{\frac{(r_1 - r_2)}{\alpha r_1 + (1-\alpha)r_2}\right\}^2 < 0$$

式（33）是存在内点解的充分必要条件。因为

$$\frac{r_1 - r_2}{\alpha r_1 + (1-\alpha)r_2}$$

---

14 $V(W_0)$ 可以由下述方法直接计算出来：如果 s 是一个常数，r^t（依照惯例 $r^0 \equiv 1$）的一个实现值所带来的效用为：

$$V(W_0) = \ln(1-s)W_0 + (1-\delta)\ln(1-s)W_0 s r^1 + (1-\delta)^2\ln(1-s)W_0 s^2 r^1 r^2 + \cdots$$

$$= \sum_{i=0}^{\infty}(1-\delta)^i \ln(1-s)W_0 s^i \Pi_{j=0}^{i} r^i$$

$$= \sum_{i=0}^{\infty}(1-\delta)^i \ln(1-s)W_0 + \sum_{i=0}^{\infty}(1-\delta)^i i\ln s + \sum_{i=0}^{\infty}(1-\delta)^i \ln\Pi_{j=0}^{i} r^j$$ 这样，

$$EV(W_0) = \frac{1}{\delta}\ln((1-s)W_0) + \frac{(1-\delta)}{\delta^2}\ln s + \frac{(1-\delta)}{\delta^2}E\ln r$$

通过下式我们可以衡量上式中的第二项，

$$f(x) = \frac{1}{1-x} = \sum_{i=0}^{\infty} x^i, f'(x) = \sum_{i=0}^{\infty} i x^i = \frac{1}{(1-x)^2}$$

因为 r^j 独立同分布，所以 $E\ln\Pi_{j=0}^{i}(r^j) = iE\ln r$。这样，

$$E\sum_{i=0}^{\infty}(1-\delta)\ln\prod_{j=0}^{i} r^j = E\sum_{i=0}^{\infty}(1-\delta)^i i\ln r = \frac{1-\delta}{\delta^2}E\ln r$$

15 若通过选择 s 来最大化式（31），那么有 $1/(1-\delta) = (1-\delta)/\delta^2 s$——这引出了式（32）。

是 r_1 的凹函数，r_1 易变性的增加总是使得投资组合中 r_2 的份额增加，而储蓄率不发生变化。

4. 企业的生产问题[16]

考虑一个企业，它下一期产出 Q 是不确定的（比如，一个公共事业企业，它必须在固定的价格上满足所有的需求）。它希望最小化生产 Q 的期望成本。生产 Q 的技术是一个两要素的、凹的生产函数 $Q=P(K, L)$，其中 K 是短期内不能变化的要素（如资本），L 是可变要素（如劳动力）。如果 Q 变得更加易变，期望成本会如何变化？如果 r 是资本的成本，w 是劳动力的成本，期望成本就由下式给出：

$$E[rK+wL(K,Q)] = rK+wE[L(K,Q)] \tag{34}$$

此处的 $L(K, Q)$ 是给定 Q 时，在资本 K 水平下所需要的劳动力。因为 F 是凹的，很容易证明给定资本 K，$L(K, Q)$ 是 Q 的凸函数。这样，Q 的易变性增加总是导致期望成本的增加。

一个稍微有些难度的问题是：最优水平的 K 如何确定？不出意外，答案取决于 K 和 L 之间的替代弹性。我们选择 K 以最小化期望成本。从式（34）可看出，一阶条件可以写为：

$$\frac{r}{w} = E\frac{\partial L(Q,K)}{\partial K}$$

即要素—价格比率必须等于平均边际替代率。我们假设生产函数具有不变替代弹性。这样，[17]

$$Q = (\delta K^{\rho} + (1-\delta)L^{\rho})^{1/\rho}$$

$$\frac{\partial L}{\partial K} = \frac{\delta}{1-\delta}\left(\frac{K}{L}\right)^{\rho-1} = \left(\frac{\delta}{1-\delta}\right)\left(\frac{Q^{\rho}-\delta K^{\rho}}{1-\delta}\right)^{(1-\rho)/\rho}K^{\rho-1}$$

$$\frac{\partial^2(\partial L/\partial K)}{\partial Q^2} = \left[\frac{\delta K^{\rho-1}}{(1-\delta)^{1/\rho}}(1-\rho)Q^{\rho-2}(Q^{\rho}-\delta K^{\rho})^{(1-3\rho)/\rho}\right] \times(-\rho Q^{\rho}+(1-\rho)\delta K^{\rho})$$

凸性的一个充分条件是 $\rho \leqslant 0$，即替代弹性小于或等于 1。这样，如果替代弹性小于或等于 1，当 Q 的易变性增加时，最优水平的 K 也会增加。

在某些生产函数下，Q 的易变性增加时，K 有可能减少；为了证明这一点，我们考虑一个极端的例子：不变替代弹性（弹性为无穷）的生产函数，

16 对此问题的一个更完整的讨论，请参见［参考文献 10］。

17 如果 $L>0$，否则 $\partial L/\partial K=0$。

$$Q = bK + aL$$

如果资本存量是 K，期望成本为：

$$rK + \frac{w}{a}\int_{bK}^{\infty}(Q - bK)dG(Q)$$

其中，$G(Q)$ 是 Q 的分布函数。期望成本最小化要求（存在内点解时）：

$$r - \frac{wb}{a}(1 - G(bK)) = 0$$

因此

$$K = \frac{G^{-1}(1 - (ar/wb))}{b}$$

此处 K 是增加还是减少仅取决于 $G^{-1}(1 - (ar/wb))$ 是增加还是减少（见图2），或者，等价的，Q 大于 bK（原资本存量的"生产能力"）的概率增加还是减少。

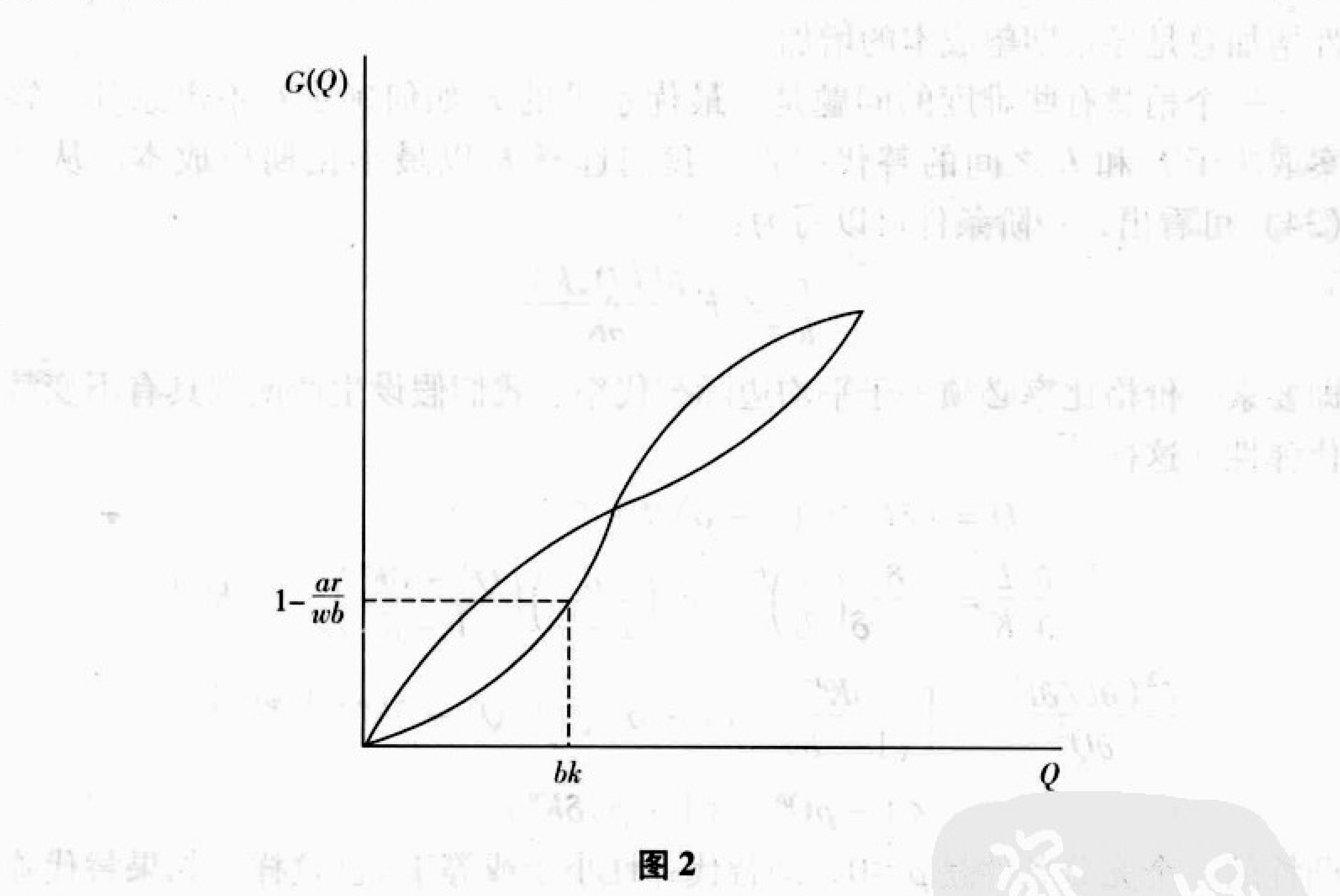


图 2

5. 多阶段计划问题[18]

考虑一个简单经济，其中的最终消费品是由劳动力和一个中间产品 y 生产的：

18 这一问题是由 M. Weitzman 向我们提出的。

$$Q = P(L_2, y)$$

而 y 的生产仅需要劳动力：

$$y = M(L_1)$$

整个经济面临一个总的劳动力限制 L，这样

$$L_1 + L_2 = L$$

不存在不确定性时，最大化 Q 仅需要

$$P_1 = P_2 M'$$

假设生产 y 时存在不确定性：

$$y = M(L_1) + e$$

其中 e 的均值为零，分布函数为 F。我们希望最大化 EQ；要求：

$$E[P_1 - P_2 M'] = 0$$

如果 e 变得更加易变，劳动力在两个部门间的配置取决于下式的符号：

$$P_{122} - M' P_{222}$$

假设 P 是一个不变替代弹性的生产函数：$P = (\delta L_2^{\rho} + (1-\delta) y^{\rho})^{1/\rho}$。那么，

$$P_{122} = \frac{A}{L_2}((1-\rho)\delta L_2^{\rho} + \rho(1-\delta)y^{\rho})$$

$$P_{222} = \frac{A((\rho-2)\delta L_2^{\rho} - (1+\rho)(1-\delta)y^{\rho})}{y}$$

其中

$$A = \delta(1-\delta)(\rho-1)L_2^{\rho} y^{\rho-2}(\delta L_2^{\rho} + (1-\delta)y^{\rho})^{(1-3\rho)/\rho} < 0$$

如果 $1 \geqslant \rho \geqslant 0$，即替代弹性大于或等于 1，且 $P_{122} \leqslant 0$，$P_{222} \geqslant 0$，那么 L_2 减少，而 L_1 增加；更多的劳动力分配于生产的“较早”阶段（以生产 y）。

考虑另外一个极端的例子，其中 Q 的生产技术是一个固定系数的生产函数：$Q = \min(L_2, y)$。这样，

$$E(Q) = \int_{\infty}^{L_2 - M(L_1)} [M(L_1) + e] dF(e) + L_2(1 - F(L_2 - M(L_1)))$$

$$= \int_{\infty}^{\bar{L} - L_1 - M(L_1)} [M(L_1) + e] dF(e) + (\bar{L} - L_1)(1 - F(I - L_1 - M(L_1)))$$

因此，最大化 EQ 要求：

$$[M'(L_1) + 1] F(\bar{L} - L_1 - M(L_1)) = 1$$

因为 $M''F - f(M'+1)^2 < 0$（f 是对应于 F 的密度函数），二阶条件也得到了满足；这样，存在唯一的最优值。L_1 是增加还是减少仅取决于 $F(L - L_1 -$

$M(L_1)$）是增加还是减少，即（在原配置中）y 的约束是紧的（binding）的概率是增加还是减少；两种情况都是非常可能的。请注意，如果 y 也是由规模报酬不变的生产函数生产的

$$y = L_1$$

那么 L_1 的最优值就由下式给出：

$$F(L - 2L_1) = \frac{1}{2}$$

因此，L_1 如何变化完全取决于 e 的中值（median）是增加还是减少。

6. 竞争性企业产出水平的选择

在此前考虑的所有例子中，要对易变性增加的作用做出确定性的结论，需要满足的条件，与较早前得出的比较无风险和有风险的情形，本质上是一致的。但是，在某些问题中，比较后者所需要的条件要弱于前者。例如，在下述例子中，我们可以给出确定性的论断，哪怕一阶条件既不是凹的，也不是凸的。

考虑一个竞争性企业，它必须在今天决定明天的产出水平，而产出 Q 的价格 p 是不确定的。它希望最大化利润的期望效用 $U(\pi)$，U 为凹函数，[19] 且

$$\pi = pQ - C(Q)$$

此处的 $C(Q)$ 是凸的成本函数。最优值的一个充分必要条件是：

$$\frac{EU'p}{EU'} = C'(Q^*) \tag{35}$$

如果生产商为风险中性，或者如果 p 不存在易变性，那么利润最大化要求价格等于边际成本，$Ep = C'(\hat{Q})$。当 $EU'p/EU' \gtrless Ep$（即 $[E(U' - EU')(p - E(p))] \gtrless 0$）时，$Q^* \gtrless \hat{Q}$。但是，由于 $U'' < 0$，当 $p \gtrless E(p)$ 时，$U'(p) \gtrless U'(E(p))$；因此，

$$E[(U' - EU')(p - E(p))] = E[U' - U'(E(p))(p - E(p))] < 0$$

这样，存在不确定性时的产出总是少于确定状态性下的产出。

不出意外，对比确定性下的竞争性企业，对于该企业行为的比较统计分析有显著的不同：

（a）不存在不确定性时，对利润征收比例为 t 的税收不会影响产出。但在本例中，它会增加或者减少产出（取决于相对风险厌恶系数是递增还是

19　对不变绝对风险厌恶系数情形的讨论［参考文献6］。除非在极具限制性的条件下，对不确定性条件下的竞争企业行为应该如何描述，这还是一个有待商榷的问题。

递减）。很容易看出，dQ/dt 与 $ERU'(p-C'(Q^*))$ 符号一致；从式（35）可得，$ERU'(p-C'(Q^*))$ 的符号与 R' 一致（再一次，$R=-U''\pi/U'$，为相对风险厌恶系数）。

（b）不存在不确定性时，对总成本曲线作一个（一致的）向上平移，对产出不会有影响。在本例中，如果

$$C(Q,\tau)=C(Q)+\tau$$

那么

$$\text{当}-EU''(p-C')=EA(p-C')U'\geqslant 0\text{时},\frac{dQ}{d\tau}\geqslant 0$$

其中 $A=-U''/U'$ 是绝对风险厌恶系数。但是 $EA(p-C')\ U'$ 与 A' 的符号一致。这样，当绝对风险厌恶系数增加（或减少）时，产出会增加（或减少）。

三、概率分布的选择

我们关于前一篇文章中讨论的对易变性的定义以及关于三种方法的等价性的基本定理，可以用来证明一些一般的定理（这些定理与个人必须在分布函数的一个可能集中选择分布函数的情形相关）。下述的例子表述了我们如何运用这些基本定理和定义。

1. 分散投资定理（Diversification theorem）[20]

假设个人可以购买两种证券[21]，他们下一期的（每美元投资的）价值是独立同分布的。他应如何配置他的初始财富？换句话说，他应如何选择 b 来最大化[22]

$$EU(W)=EU((be_1+(1-b)e_2)W_0)$$

其中 U 是凹函数。我们将证明 b 应固定在 $\frac{1}{2}$，且独立于效用函数。我们可以得到下式：

$$y_b=(be_1+(1-b)e_2)W_0=y_{1/2}-(b-\frac{1}{2})(e_1-e_2)W_0$$

请注意，

$$E(e_1-e_2\mid y_{1/2})=0$$

即给定两个独立同分布的随机变量之和为一个特定的数值，它们的期望值之

20 ［参考文献 12］给出了另一种证明，并对此定理做了一般的讨论。

21 很容易将结论推广至存在 n 种证券的情形。

22 我们假设 $E(et)$ 存在并且有限，所以 EU 存在并且有限。

差为零。因为 y_b 的分布与 $y_{1/2}$ 加上一个随机变量（条件于 $y_{1/2}$ 的期望为零）的分布相同，根据前一篇文章中的定理2，我们可以得出，对于所有 $b \neq \frac{1}{2}$，y_b 比 $y_{1/2}$ 更易变，即对于效用函数为凹的个人来说，$y_{1/2}$ 比 y_b 更好。

2. 劳—布莱克威尔（Rao-Blackwell）定理[23]

假设未知变量 θ 决定的分布产生了一个随机变量样本，$x = (x_1, \cdots, x_n)$。估计 θ 的一种方法是利用从样本 x 到实轴的映射 $d(x)$。通常，统计学家希望找到一个能够使得凸的损失函数 $L(d(x))$ 期望值最小的估计程序。劳—布莱克威尔定理说明，对于任一估计量 $d(x)$，以及任一凸函数 L，如果存在一个 θ 的充分统计量 T，那么存在一个估计量 $d^*(x)$，它至少不差于 $d(x)$（从 $EL(d^*(x)) \leqslant EL(d(x))$ 来讲）。

为证明这一点，从第一部分的定理2可得，其充分必要条件是找到满足 $d^*(x) \leqslant_a d(x)$ 的 $d^*(x)$。

对于所有的 T，令 $d^*(x) = E(d(x) \mid T)$。考虑一个随机变量 z，满足 $d(x) = d^*(x) + z$。显然，正如我们要证明的，$E(z \mid T, x) = E(z \mid d^*(x)) = 0$ 且 $d^*(x) \leqslant_a d(x)$。

参考文献

1. K. J. Arrow, "Aspects of a Theory of Risk Bearing," Yrjö Jahnsson Lectures, Helsinki, 1965.
2. A. B. Atkinson, On the measurement of inequality, *J. Econ. Theory*.**2** (1970), 244–263.
3. T. S. Ferguson, "Mathematical Statistics," Academic Press, New York/London, 1967.
4. F. H. Hahn, Savings under uncertainty, *Rev. Econ. Studies* **37** (1970), 21–24.
5. D. Levhari and T. N. Srinivasan, Optimal savings under uncertainty, *Rev. Econ. Studies* **36** (1969), 153–164.
6. J. J. McCall, Competitive production for constant risk utility functions, *Rev. Econ. Studies* **31** (1964), 417–419.
7. J. A., Mirrlees, "Optimum Accumulation under Uncertainty," December 1965, Cambridge, unpublished.
8. E. S. Phelps, The accumulation of risky capital: A sequential utility analysis, *Econometrica* **30** (1962), 729–743.
9. J. W. Pratt, Risk aversion in the small and in the large, *Econometrica* **32** (1964),

23　这里的定义和记号参考了 Ferguson［参考文献3］。

122–136.
10. M. Rothschild, "Changing Demand: Its Costs and Consequences," paper presented to the Econometric Soc., Evanston, December, 1968.
11. M. Rothschild, and J. E. Stiglitz, Increasing risk I: A definition, *J. Econ. Theory* **2** (1970), 225–243.
12. P. A. Samuelson, General proof that diversification pays, *J. Finan. Quant. Anal.* **2** (1967), 1–13.
13. J. E. Stiglitz, The effects of wealth, income and capital gains taxation on risk taking, *Quart. J. Economics* **83** (1969), 263–283.
14. J. E. Stiglitz, A consumption oriented theory of the demand for financial assets and the term structure of interest rates, *Rev. Econ. Studies*, **37**, 321–351.
15. James Tobin, The theory of portfolio selection, *in* "The Theory of Interest Rates" (E. Hahn and F. Brechling, Eds.), Macmillan, London, 1965.
16. James Tobin, Comment on Borch and Feldstein, *Rev. Econ. Studies* **36** (1969), 13–14.

投资者偏好和资产回报的结构与投资组合配置的分离性：对共同基金纯理论的贡献*

一、引言

10多年前，托宾［参考文献5］证明了在一定条件下，投资者选择投资组合配置的过程可被看作是两阶段的：投资者首先决定以多大比例去购买风险资产，然后再决定如何将他的总投资在风险资产和安全资产之间进行分配。这类型的“分离”是投资者选择最优投资组合的一种特殊情形，更一般的特征是在给定的市场上，存在着 n 种不同的资产，但是与投资者的决策有关的所有市场机会都可以被一组 m（$<n$）种“共同基金”（即可用资产加权和为1的线性组合）提供。

我们之所以对这种特征感兴趣，至少有四个原因：（i）凯恩斯学派的传统宏观经济模型假定存在这样的分离性。（ii）当这样的分离性存在时，实现纯粹交换经济下的帕累托最优，并不需要完全的阿罗—德布鲁（Arrow-Debreu）证券的集合。当然，也存在重要的理由（比如说交易成本）让人相信一个完全的阿罗—德布鲁市场并不存在。（iii）现代投资理论的很多结果的关键假设就是存在安全资产和唯一的一种风险资产，或等价地，由一种风险资产构成的共同基金。[1]（iv）分离性是经济学家长期以来感兴趣的加总特征的一种特殊扩展。[2]

---

* “The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Fund”, with David Cass, Journal of Economiu Theory, 1, June 1970, pp. 120—160. 这篇论文的研究是在美国国家科学基金会和福特基金会的资助下进行的。我们得益于很多人的评论，但特别要感谢高曼（W. M. Gorman）和萨甘（J. D. Sargan）指出了我们最初的论证中忽略的一个地方。

1 例如，可参见阿罗［参考文献1］和卡斯—斯蒂格利茨［参考文献2］，分别分析了存在和不存在这样的假设时，财富的变化对于投资组合配置的效果。

2 传统的加总理论关心这样一个问题，何时能够把方程 $F(x_{11},\cdots,x_{1n_1},x_{21},\cdots,x_{mn_m})$ 写作是 $f[X_1(x_{11},\cdots,x_{1n_1}),\cdots,X_m(x_{m1},\cdots,x_{mn_m})]$，其中当 $i=1,\cdots,m$ 时 X_i 是加总量?在我们的问题中，我们寻找条件使我们能够把方程写作 $f[X_1(x_{11},\cdots,x_{mn_m}),\cdots,X_k(x_{11},\cdots,x_{mn_m})]$，其中当 $i=1,\cdots,k<\sum n_j$ 时 X_i 是初始变量加和得到的加总量。

这篇文章有双重目的：首先是找到这样的共同基金得以形成的充分必要条件（对应于投资者的偏好），然后更加限制性地，寻找能够分离为风险资产和安全资产的条件。我们的结果是具有启发性，因为它们不仅证明了这些条件的限制有多么强，而且它们分离出“货币”（安全资产通常的称谓）发挥的特殊作用。

二、问题的正式阐述

个人投资者有初始财产 W_0。它可以用来购买 n 种不同的风险资产或证券。如果投资者在第 i 种证券的投资额为 Z_i，那么他的预算约束可以简单地表述为：

$$\sum_{i=1}^{n} Z_i \leqslant W_0 \tag{1}$$

投资者期望状态 θ 发生时，第 i 种证券的毛回报（对每一美元投资）为 $\rho_{i\theta} > 0$。因此对于任何给定的投资组合选择 $Z_i, i = 1, \cdots, n, W_\theta = \sum_{i=1}^{n} Z_i \rho_{i\theta}$ 就是在下一期，如果状态 θ 发生，投资者的财富。投资者选择投资组合，以最大化下一期财富的期望效用：

$$EU(W_\theta) = EU\left(\sum_{i=1}^{n} Z_i \rho_{i\theta}\right) \tag{2}$$

同时满足预算约束式（1）。我们假定 U 是二阶可微且严格凹的，即 $U''(W) < 0$。[3] 期望效用函数最大化的充分必要条件，除了式（1）外，还包括[4]

$$EU'(W_\theta)\rho_{i\theta} = \lambda, i = 1, \cdots, n \tag{3}$$

其中 λ 是初始财富的边际效用。

---

3 当然，为了使问题有意义，我们必须同时假定 $U'(0) > 0$。然而我们并没有事先排除餍足的可能性，即 $U'(\overline{W}) = 0$ 对于某一的 $\overline{W} < \infty$。这是可能的情形，例如在投资组合理论中经常假设效用函数是二次型。

4 在关于证券市场的准则和期望回报结构的两个进一步假设下：（i）卖空是允许的；（ii）任何证券的回报都不能够被其他证券在加权和（权重可能为负）为 1 的线性组合的回报占优。给定第一个假设，要使投资组合配置的问题有意义，第二个假设是必需的；例如，如果投资者相信某一种证券会被另一种占优，便能够卖空第一种，那么他就能够从他的投资组合中获得在每一种状态或是或然状态下任意多的回报。显然允许卖空关系到投资组合行为的重要结论，但这样的假设对我们的目的而言基本没有坏处；如果不允许卖空，可以直接将我们的结论进行必要的修改。

有时我们表示为第 i 种证券 Z_i 占初始财富的比例

$$a_i = \frac{Z_i}{W_0} \tag{4}$$

而不是分配到第 i 种证券的初始财富量 Z_i 会有帮助。此时由 a_i 表示的预算约束变为

$$\sum_{i=1}^{n} a_i \leqslant 1 \tag{1'}$$

共同基金是指为任何一种投资者可用的初始证券的加权和为1的固定线性组合构成的证券。因此共同基金的回报 μ_θ 有如下形式：

$$\mu_\theta = \sum_{i=1}^{n} \delta_i \rho_{i\theta} \quad 其中 \sum_{i=1}^{n} \delta_i = 1 \tag{5}$$

那么，我们称一类效用函数[5]具有广义分离性，当且仅当对任何 n 种初始证券的集合，存在着 m（$<n$）种共同基金

$$\mu_{j\theta} = \sum_{i=1}^{n} \delta_{ij} \rho_{i\theta}, j = 1, \cdots, m \tag{6}$$

使得具有这种效用函数的投资者的最优配置，可以被简单地表示为共同基金的配置：

$$a_i = \sum_{j=1}^{m} a'_j \delta_{ij} \quad 其中 \sum_{j=1}^{m} a'_j = 1, i = 1, \cdots, n \tag{7}$$

这种性质的等价描述是我们能够找到 m 种共同基金，使得任何具有该类效用函数的投资者，在任何财富水平上，面对这些共同基金提供的市场机会能实现的期望效用最大化值与面对 n 种初始证券提供的市场机会能实现的值是一样的。[6] 注意，一般来说定义共同基金的权重（共同基金的数量更是如此）将同时取决于效用函数的类型和初始证券的回报。

对广义分离的几何解释如下：假定不存在餍足，我们只需要考虑式（1）取等号的情形（其他情形类似）。那么投资到每种证券的初始财富的比例 a_i 所表示的可行投资集，恰好是一个包括单纯形的 $n-1$ 维的超平面（如果不允许卖空，可行投资集将是单位单形（Unit Simplex））。广义分离意味着，最优投资集（当初始财富水平变化时），用 a_i 表示是可行集包含的 $n-2$

5　"这一类效用函数"的一个例子是凹的，二次型函数 $U'(W) = \alpha + \beta W + \gamma W^2$，其中 $\beta > 0$，$\gamma < 0$。

6　因此，式（7）中 a'_j 代表了投资组合投资在第 j 种共同基金中的最优比例。

维（或更小）的超平面。$n=3$ 的可能情形如图1所示。其中可行集为平面 P，对于显示出广义分离的效用函数，最优集是平面上的一条直线，比如 AB，或者平面上的一个点，比如 C。对于前者，如果两种共同基金是按 AB 上的任何两个点对应的三种资产的比例形成的，那么投资者通过购买这两种共同基金便得到他的最优配置。类似地，对于后者，投资者只需获得按 C 给出的资产比例形成的共同基金。

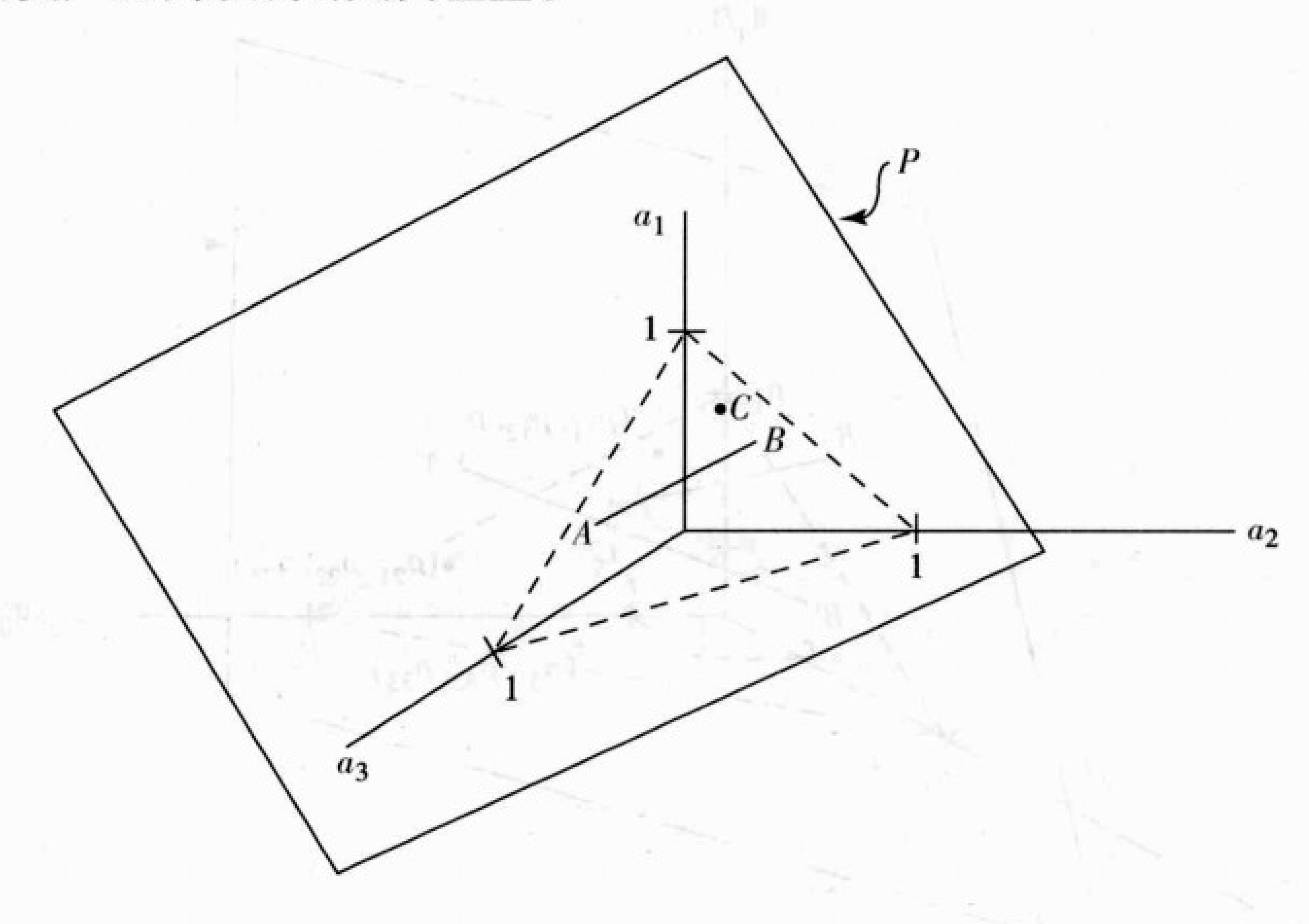


图1

为了方便起见，我们说广义分离达到便是指某一情形中效用函数具有广义分离性。当共同基金的数目是两个时（$n=2$），我们称该类型效用函数显示出分离性或说存在分离。我们主要分析这种更有限制性的性质和情形。

广义分离性和分离性都是关于效用函数的陈述，而且效用函数是独立于可用资产的回报结构的。我们也会关注一种特殊的情形，它具有这样的回报结构：或者存在货币，也就是在任何的自然状态下都具有相同回报的资产，或者有可能形成一种共同基金作为货币。在这种情形下，如果：(i) 分离达到；(ii) 形成的两种基金中的一种总会是货币，那么我们称货币分离达到。

分离与货币分离的区别如图2所示。假设有三种自然状态，同时也只有三种证券（我们在附录Ⅰ清楚地证明了，要使这个例子当卖空被允许时有意义，除了别的条件外，还要求形成一种共同基金作为货币）。由初始财富在每种状态下的回报 $W_\theta/W_0=\sum_{i=1}^{n}a_i\rho_{i\theta}$ 表示的可行投资集由平面 P 给出，它穿

过三个点（ρ_{i1}，ρ_{i2}，ρ_{i3}），分别对应初始财富全部投入到每一种可用资产得到的回报。如果分离达到，投资者将选择投资沿着这个平面上的一条直线，比如 AB，而货币分离达到时，这条直线将会通过这样一点，在这一点上

$$W_1/W_0 = W_2/W_0 = W_3/W_0 = r$$

比如 $A'B'$。

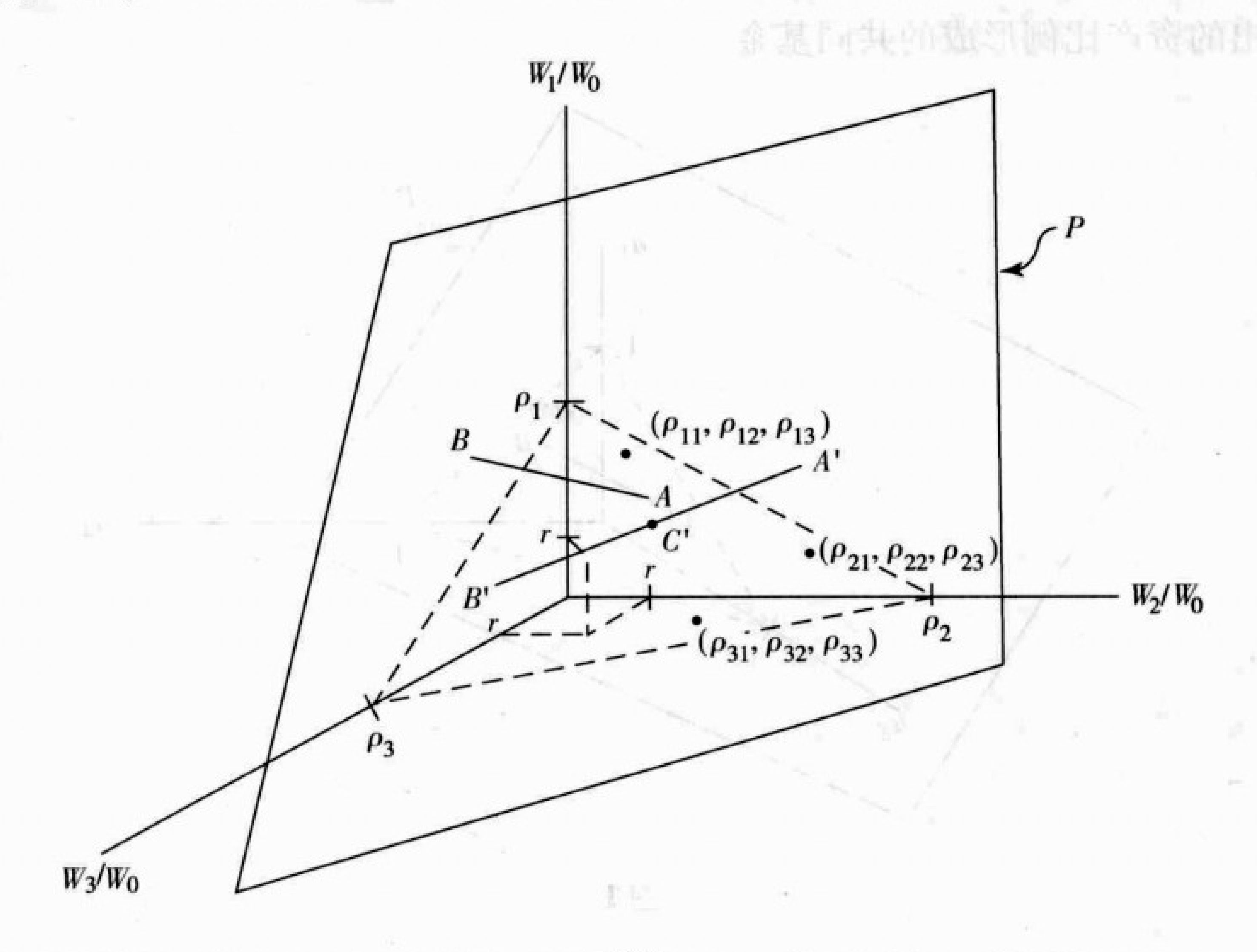


图 2

三、存在阿罗—德布鲁证券时的分离

我们将可用资产回报的一种非常特殊的结构作为分析具有分离性的效用函数的起点。这种方法的基本原理如下：我们把具有分离性的效用函数简称为可行的效用函数，显然如果它能够在一般的情形下可行，在回报结构特别给定时也必然可行。此外，通过限制回报的结构，我们发现仅有相对来说少数的效用函数可行，那么，验证它们之中哪些在没有回报结构的限制时可行，便是简单的事情。为了实现上述过程，我们首先只考虑能够为每一种或有状态提供索取权的证券（这里或有状态即为自然的状态）。

为了更精确地阐明我们所说的“为每一种或有状态提供索取权”的含义：假定存在着 n 种自然的状态，同时有 n 种证券，同时证券 i 的回报（对

于每一美元投资）为：当状态 $\theta = i$ 出现时，$\rho_{i\theta} = \rho_i > 0$，否则 $\rho_{i\theta} = 0$。那么投资者在状态 i 的财富仅取决于他在证券 i 上的投资，或者换一种方式说，投资者可以通过配置他的投资组合以保障每一种或有状态。（当然要满足他总的预算约束）我们把这样的证券称为阿罗—德布鲁证券（附录Ⅰ证明了，在适当的限制下，当存在着和自然状态一样多的证券时，任何给定的证券集，都能找到一个等价的阿罗—德布鲁证券集）。

这种特殊的回报结构的重要推论即期望效用函数最大化的一阶条件具有特殊形式

$$U'(W_i)\pi_i\rho_i = \lambda,\ i = 1,\cdots,n \tag{3'}$$

其中 $W_i = Z_i\rho_i$ = 下一期状态 i 发生时的财富，π_i = 状态 i 发生的概率（因此满足 $\pi_i > 0$ 和 $\sum_{i=1}^{n}\pi_i = 1$）。假定分离达到与参数 ρ_i 和 π_i 的值无关，通过分析式（3′）可以得到下面的基本定理：

定理3.1：给定阿罗—德布鲁证券，分离的充分必要条件是边际效用满足

$$AU'(W)^{\alpha} + BU'(W)^{\beta} = W \tag{8}$$

或者

$$U'(W)^{\alpha}(A + B\log U'(W)) = W \tag{9}$$

注释1：这些方程包含的一些特殊情形在以前对不确定性下行为的讨论中被广泛使用，我们接下来也会特别地关注这些情形：

(i) 如果在式（8）中，$\alpha = 0$，我们将得到边际效用的方程为如下形式

$$U'(W) = (a + bW)^{c} \tag{10}$$

其中 $b = 1/B$，$a = -bA$，$c = 1/\beta$。式（10）的两种特殊的子情况是人们已熟知的，一种是同时要求 $A = a = 0$（或者等价地，在式（9）中，$B = 0$），$U'(W) = bW^{c}$，常相对风险厌恶（constant relative risk aversion）效用函数，[7]

$$\tag{10'}$$

---

7 之所以这样称呼式（10′），是因为阿罗—普拉特［参考文献1、4］相对风险厌恶的度量，$-U''(W)W/U'(W)$ 是常数（$= -c$），式（11）是因为它们的绝对风险厌恶的度量，$-U''(W)/U'(W)$，是常数（$= -b$）。

$Ax^{\alpha} + Bx^{\beta} = W$	$x > 0$ 且 $Ax^{\alpha} + Bx^{\beta} > 0$	$\alpha Ax^{\alpha} + \beta Bx^{\beta} < 0$
$x^{\alpha}(A + B\log x) = W$	$B > 0$ 且 $x > e^{-A/B}$	$\alpha < 0$ 且 $x > e^{-(A/B+1/\alpha)}$
	$B < 0$ 且 $0 < x < e^{-A/B}$	$\alpha < 0$ 或者 $\alpha > 0$ 且 $x < e^{-(A/B+1/\alpha)}$
$x = (a + bW)^{c}$	$b > 0$ 且 $W \geqslant \max[0, -(a/b)]$	$c < 0$
	$a > 0, b < 0$ 且 $0 \leqslant W < -(a/b)$	$c > 0$
$x = ae^{bW}$	$a > 0$ 且 $0 \leqslant W$	$b < 0$

另一种是同时要求 $\beta = c = 1$，

$$U'(W) = a + bW, \text{二次型效用函数} \tag{10''}$$

(ii) 如果，在式（9）中，$\alpha = 0$，我们将得到边际效用的方程为如下形式

$U'(W) = ae^{bW}$，常绝对风险厌恶（constant absolute risk aversion）效用函数　(11)

其中 $b = 1/B$，$a = e^{-bB}$。

注释2：一般来说，式（8）和式（9）都不能够通过变换导出效用函数 $U(W)$ 的解析式。然而，对于上面提及的特殊情形，以及式（8）的另一类特殊情形（即那些 α 和 β，A 和 B 是复共轭而且它们实部为0，可参见附录Ⅱ的第一个脚注），这样的解析式是可以得到的。

由于计算的细节很简单，我们将其忽略。值得提及的是，即使式（8）和式（9）无法通过求逆来获得 U' 用 W 表示的显式解（推出解析式的关键步骤），我们仍然有可能得到 U 的隐含形式，通过对式（8）和式（9）中的每一项进行求逆和积分，可得

$$U(W) = \max_{\substack{W_1 + W_2 = W \\ W_1, W_2 \geq 0}} A'W_1^{\alpha'} + B'W_2^{\beta'} + C' \tag{12}$$

或者

$$U(W) = \max_{\substack{W_1 + W_2 = W \\ W_1, W_2 \geq 0}} A'W_1^{\alpha} + B'\int_{C'}^{W_2} f(z)\,dz \tag{13}$$

其中 f 是函数 $z = x^{\alpha}\log x$ 的反函数。[8]

注释3：当然，并不是所有满足式（8）或式（9）的函数都能生成一个凹的效用函数。为了完整，我们列出参数的限制以及式（8）至式（11）中函数的定义域和值域的限制以使得当 $0 \leq \underline{W} < W < \overline{W} < \infty$ 时，$U'(W) > 0$ 且 $U''(W) < 0$：

同时，要进一步

如果 $U' \equiv x$ 满足　要使 $U' > 0$，要求　有 $U'' < 0$，还要求

8　对于这样一种形式的问题 $\max g_1(W_1) + g_2(W_2)$，满足 $W_1 + W_2 = W$，同时 $W_1, W_2 \geq 0$ 的目标函数值，比如说 $U(W)$（例如式(12)），可以被这样一个方程刻画 $g'^{-1}_1[U'(W)] + g'^{-1}_2[U'(W)] = W$（例如式（8）），它来自于一阶条件

$$\begin{cases} W_1 + W_2 = W \\ g'_i(W_i) \leq U'(W) \quad \text{当 } W_i > 0 \text{ 时取等号} \end{cases}$$

例如，当方程 g_i 是递增，凹的而且满足 $\lim_{W_i \to 0} g_i'(W_i) = \infty$。考虑到这一点，可以直接推出式（8）和式（12）或者式（9）和式（13）等价的条件。

定理3.1的证明：证明将分为几个步骤。为了叙述方便，讨论将集中在$n=3$的情形；然而证明却是非常一般化的。

（1）分离性等价于对证券的需求函数具有如下形式

$$Z_i = A_i W_0 + B_i h(W_0), i = 1,2,3 \tag{14}$$

其中系数A_i和B_i满足$\sum_{i=1}^{3} A_i = 1, \sum_{i=1}^{3} B_i = 0$。

这一点可以直接由分离的定义给出（正如式（14）仅是由$\sum_{i=1}^{3} a_i = 1$刻画的平面上的一条直线的参数表达）。注意到这一事实，即只存在两种共同基金，且投资者对证券的需求可以被写成这些共同基金的线性组合（独立于初始财富），这意味着：

$$\begin{aligned} Z_i &= [a'(W_0)\delta_{i1} + (1 - a'(W_0))\delta_{i2}] W_0 \\ &= \delta_{i2} W_0 + (\delta_{i1} + \delta_{i2}) a'(W_0) W_0 \end{aligned}$$

其中$a'(W_0)$是分配到第一种共同基金上的投资组合的比例（参考式（6）和式（7））。[9]

（2）使得需求函数具有式（14）的形式的必要条件是$U'^{-1} \equiv G \neq 0$满足函数方程

$$G(xy) = \bar{f}(x)\tilde{f}(y) + \bar{g}(x)\tilde{g}(y) \tag{15}$$

其中$\bar{f}$，$\tilde{f}$，$\bar{g}$和$\tilde{g}$满足（a）当$0 \leqslant \underline{x} = \max[0, U'(\infty)] < x < U'(0) = \bar{x} < \infty$时，$G(x)$，$\tilde{f}(x)$和$\tilde{g}(x)$有定义而且可微，而且（b）当$x > 0$时$\bar{f}(x)$和$\tilde{g}(x)$有定义而且可微。

对于一个具有分离性的效用函数，可以考虑这样的情形：存在三种阿罗—德布鲁证券以概率$\tilde{\pi}_i$产生回报$\tilde{\rho}_i$。为了简单起见，我们把这样的回报和概率的结构称为一个阿罗—德布鲁市场（或一般地，任何回报和概率的结构都简称做一个市场）。对于这样一个特别的市场，式（1）和式（3′）可以被写成（不存在餍足可能性时，对于任何初始财富水平）

$$\sum_{i=1}^{3} Z_i = \sum_{i=1}^{3} \frac{W_i}{\tilde{\rho}_i} = W_0 \tag{16}$$

$$U'(W_i)\tilde{\pi}_i\tilde{\rho}_i = \lambda, i = 1,2,3$$

9　显然，式（14）表示的分离，不管怎样的回报结构都成立，比如说，不管证券是否是阿罗—德布鲁证券。

利用风险厌恶的假设 $U''<0$，我们首先可以认为这四个方程把 λ 定义为 W_0 的递减函数，或者更符合我们的目的：W_0 是 λ 的递减函数

$$W_0 = \tilde{\varphi}(\lambda) \text{ 其中} \tilde{\varphi}' < 0 \tag{17}$$

其次，我们能够将最后三个方程求逆，从而得到需求函数的另一种形式

$$Z_i = \frac{1}{\tilde{\rho}_i} U'^{-1}\left(\frac{\lambda}{\tilde{\pi}_i \tilde{\rho}_i}\right), \quad i = 1,2,3 \tag{18}$$

将式（14）和式（18）取等，再利用式（17），我们最后得到

$$\begin{aligned} U'^{-1}\left(\frac{\lambda}{\tilde{\pi}_i \tilde{\rho}_i}\right) &= A_i \tilde{\rho}_i \tilde{\varphi}(\lambda) + B_i \tilde{\rho}_i h[\tilde{\varphi}(\lambda)] \\ &= \tilde{A}_i \tilde{F}(\lambda) + \tilde{B}_i \tilde{G}(\lambda), i = 1,2,3 \end{aligned} \tag{19}$$

式（19）非常接近函数式（15），只要令 $x = (1/\tilde{\pi}_i \tilde{\rho}_i)$ 和 $y \equiv \lambda$。[10]唯一的障碍是我们还没有证明对于任何给定的函数方程，$\tilde{F}$ 和 $\tilde{G}$ 独立于特定的阿罗—德布鲁市场而且 $\tilde{A}_i$ 和 $\tilde{B}_i$ 仅取决于 $\tilde{\pi}_i \tilde{\rho}_i$。这一点不难通过考虑另一个阿罗—德布鲁市场来证明，这个市场用双重波浪号表示，可有

$$\begin{aligned} \approx{\pi}_i \approx{\rho}_i &= \tilde{\pi}_i \tilde{\rho}_i, \quad i = 1,2 \\ &\neq \tilde{\pi}_i \tilde{\rho}_i, \quad i = 3 \end{aligned} \tag{20}$$

对于这个市场，可以重复得到式（19）的证明过程，可得

$$U'^{-1}\left(\frac{\lambda}{\approx{\pi}_i \approx{\rho}_i}\right) = \approx{A}_i \approx{F}(\lambda) + \approx{B}_i \approx{G}(\lambda), \quad i = 1,2,3 \tag{21}$$

此时，通过式（20）的假设，由式（19）和式（21）开始的两个方程，可得

$$U'^{-1}\left(\frac{\lambda}{\approx{\pi}_i \approx{\rho}_i}\right) = \approx{A}_i \approx{F}(\lambda) + \approx{B}_i \approx{G}(\lambda) = \tilde{A}_i \tilde{F}(\lambda) + \tilde{B}_i \tilde{G}(\lambda), i = 1,2 \tag{22}$$

通过假设 $\approx{A}_1 \approx{B}_2 - \approx{A}_2 \approx{B}_1 \neq 0$，我们可以解出 $\approx{F}$ 和 $\approx{G}$，用 $\tilde{F}$ 和 $\tilde{G}$ 表示即为，

10　事实上，如果我们假定一个连续统的而不是有限个数的阿罗—德布鲁证券，类似式（19）的方程可以直接推出式（15）。

$$\begin{aligned}\tilde{\tilde{F}}(\lambda) &= C_1\tilde{F}(\lambda) + C_2\tilde{G}(\lambda)\\ \tilde{\tilde{G}}(\lambda) &= C_3\tilde{F}(\lambda) + C_4\tilde{G}(\lambda)\end{aligned} \tag{23}$$

因此，我们可以把式（21）中第三个方程也用 $\tilde{F}$ 和 $\tilde{G}$ 表示：

$$U'^{-1}\left(\frac{\lambda}{\tilde{\tilde{\pi}}_3\tilde{\tilde{\rho}}_3}\right) = A_3\tilde{F}(\lambda) + B_3\tilde{G}(\lambda) \tag{24}$$

其中

$$A_3 = C_1\tilde{\tilde{A}}_3 + C_3\tilde{\tilde{A}}_3, B_3 = C_2\tilde{\tilde{B}}_3 + C_4\tilde{\tilde{B}}_3$$

但是，式（24）意味着，如果 $\tilde{A}_1\tilde{\tilde{B}}_2 - \tilde{A}_2\tilde{\tilde{B}}_1 \neq 0$ 对于式（20）刻画的每一个市场都成立，我们便证明了 U'^{-1} 将满足式（15）。进一步，由于初始参数 $\tilde{\rho}_i$ 和 $\tilde{\pi}_i$ 基本上是任意给定的正数，如果附加的假设并不是对所有的参数结构都成立，U' 将会有式（10）的形式，那么 U'^{-1} 仍旧要满足式（15）。

这里我们只证明后一个命题的第一部分，第二部分将作为一个更一般的命题的特殊情形在附录Ⅱ中证明（情形 1 的第二段）。假设对于任何一种参数选择（为了简单起见，我们略去波浪号）ρ_i 和 π_i，$i=1, 2$，满足 $m \leqslant (\pi_2\rho_2/\pi_1\rho_1) \leqslant M$，且 $0<m<1<M<\infty$，都存在另一对参数使得在需求式（14）中的系数满足 $A_1B_2 - A_2B_1 = 0$。这意味着不管初始财富 W_0 是多少，来自证券 1 和证券 2 的收入都将有恒定的比例 C，或者用一阶条件式（3′）表示为：

$$U'(W_1)\pi_1\rho_1 = U'(CW_1)\pi_2\rho_2 \tag{25}$$

其中 C 明显是在某一值域 $n \leqslant C \leqslant N$，$0<n<1<N<\infty$ 上关于 $\pi_2\rho_2/\pi_1\rho_1$ 的增函数（但这个值域一般取决于所有的参数 ρ_i 和 π_i，以及 U' 的形式）。然而式（25）可推出

$$\frac{U''(W_1)W_1}{U'(W_1)} = \frac{U''(CW_1)CW_1}{U'(CW_1)} \tag{26}$$

这要在 $n \leqslant C \leqslant N$ 上都成立，当且仅当 U' 具有常数弹性，也就是具有式（10′）的形式。

式（15）中的函数的可微性直接来自于 U 是二次可微的假设下对式（24）的推导，而这些函数的定义区间仅仅由一阶条件式（16）以及阿罗—德布鲁市场结构的假设所决定。

（3）在附录Ⅱ中我们会证明式（15）只有如下解：

$$\bar{f}(x) = \frac{\tilde{f}(x)}{A} = x^{\alpha}, \bar{g}(x) = \frac{\tilde{g}(x)}{B} = x^{\beta} \text{ 和 } G(x) = Ax^{\alpha} + Bx^{\beta} \tag{27}$$

以及

$$\bar{f}(x) = \tilde{g}(x) = x^{\alpha}\left(\frac{A}{2} + B\log x\right)$$

$$\tilde{f}(x) = \bar{g}(x) = x^{\alpha} \text{ 和 } G(x) = x^{\alpha}(A + B\log x) \quad (28)$$

式（27）和式（28）得到的恰好是式（8）至式（9）中所指的U'应满足的条件，我们因而证明了必要性。由于证明充分性（即证明式（8）至式（9）结合式（3′）将推出式（14′）是显然的，我们把它留给感兴趣的读者去证明。

四、存在阿罗—德布鲁市场时的货币分离

在这一部分我们将研究，仍在阿罗—德布鲁市场的框架下，要使货币总能够被选作共同基金的一种，我们要对效用函数的形式做怎样进一步的限制。[11]论证的过程是相当独立的，[12]可以得到

定理4.1：给定阿罗—德布鲁证券，货币分离的充分必要条件是边际效用具有式（10）或式（11）的形式。

注释：这个阿罗—德布鲁市场的结果已经告诉我们货币分离所施加的限制是非常苛刻的。实际上，式（10）和式（11）这两类效用函数恰好是我们已知的，能够在一般的有货币的市场上推出货币分离的效用函数。

定理4.1的证明：（必要性）　考虑任何存在着三种证券的阿罗—德布鲁市场，并且假定效用函数能够得到货币分离，也就是分离存在而且两种基金中有一种是在各种状态下回报都是$r = 1\Big/\sum_{i=1}^{3}(1/\rho_i)$的证券（可以参考附录Ⅰ中的方程（Ⅰ.6））。此时另一种证券便是阿罗—德布鲁证券的加权和为1的线性组合，比如说第一种权重为δ_1，第二种相应为δ_2，而第三种为$\delta_3 = 1 - \delta_1 - \delta_2$（例如，图2中的$B'$点）。令货币上的投资额为$x$，那么有风险的共同基金上的投资额便是$W_0 - x$。那么投资者的最优配置可以表示为

$$\begin{aligned} W_1 &= \rho_1 Z_1 = rx + \delta_1\rho_1(W_0 - x) \\ W_2 &= \rho_2 Z_2 = rx + \delta_2\rho_2(W_0 - x) \quad (29) \\ W_3 &= \rho_3 Z_3 = rx + \delta_3\rho_3(W_0 - x) \end{aligned}$$

11　由附录Ⅰ的讨论可知这里的研究等于是在这样一个市场上分析货币分离，即可用证券覆盖所有的或有状态而且允许卖空（因此总存在一种共同基金作为货币）。

12　特别地，只对货币分离感兴趣的读者在这里或第六部分的第1点可找到我们关于这个问题的所有结果。

解出 $W_1 - W_3$ 以及 $W_2 - W_3$ 用 $W_0 - x$ 表示，式（29）得出货币分离的必要条件

$$\frac{W_1 - W_3}{W_2 - W_3} = \frac{\delta_1\rho_1 - \delta_3\rho_3}{\delta_2\rho_2 - \delta_3\rho_3}$$

或（微分并化简）

$$\frac{dW_1}{dW_0}(W_2 - W_3) + \frac{dW_2}{dW_0}(W_3 - W_1) + \frac{dW_3}{dW_0}(W_1 - W_2) = 0 \qquad (30)$$

但由刻画这个市场的最优配置的条件（不带波浪号的式（16）），我们得到关于 dW_i/dW_0 的额外信息，即

$$\sum_{i=1}^{3} \frac{dW_i}{dW_0}\frac{1}{\rho_i} = 1$$

$$U''(W_i)\pi_i\rho_i \frac{dW_i}{dW_0} = \frac{d\lambda}{dW_0}, \quad i = 1,2,3$$

或（显示地解出）

$$\frac{dW_i}{dW_0} = \frac{V_i}{\sum_{i=1}^{3} \frac{V_i}{\rho_i}} \qquad (31)$$

其中 $V_i = -U'(W_i)/U''(W_i)$ 普拉特绝对风险厌恶函数的倒数［参考文献4］（注意到，顺便指出，式（31）需要阿罗—德布鲁证券对于风险厌恶的投资者是奢侈品）。因此，把式（31）代入式（30）可以得到另一个货币分离的必要条件：

$$V_1(W_2 - W_3) + V_2(W_3 - W_1) + V_3(W_1 - W_3) = 0 \qquad (32)$$

此时我们可以看到，结合式（16）的条件，事实上式（32）必然对给定 W_2，W_3 时，$\max[U'(\infty), 0] < U'(W_1) < U'(0)$ 都成立，也就是说必然成立，如果我们改变 ρ_1 和 W_0 以保持 λ 为常数（这非常类似我们早先在推导函数式（15）使用过的论证）。因此，我们可以写作（去掉下标1）

$$-\frac{U'(W)}{U''(W)} = \alpha + \beta W$$

或

$$\frac{d\log U'(W)}{dW} = \frac{1}{\alpha + \beta W} \qquad (33)$$

当 $\beta \neq 0$ 时式（33）的解析解恰好为式（10），$\beta = 0$ 时为式（11）。

（充分性）要证明式（10）和式（11）能够推出阿罗—德布鲁市场上的货币分离可以直接将 U' 的这些形式代入式（16）的条件中求解（其中3被替换为 $n \geqslant 3$）：

$$U'(W) = (a + bW)^c \qquad (\text{i})$$

在这种情形，条件式（16）变为

$$\sum_{i=1}^{n}\frac{W_i}{\rho_i}=W_0 \tag{34}$$

$$(a+bW_i)^c\pi_i\rho_i=\lambda, i=1,\cdots,n$$

通过式（34）中的最后 n 个方程可以解出 W_i，把这些结果代入第一个方程同时解出 λ，那么，最终利用这个结果把 W_i 表示为用初始财富 W_0 和参数 a，b，c，ρ_i 和 π_i 的式子可得

$$W_i=rx+\delta_i\rho_i(W_0-x), i=1,\cdots,n \tag{35}$$

其中

$$r=1\Big/\sum_{i=1}^{n}\frac{1}{\rho_i}, x=-\frac{a}{b}\sum_{i=1}^{n}\frac{1}{\rho_i} \quad 而 \quad \delta_i=\frac{(\pi_i\rho_i)^{-1/c}}{\rho_i}\Big/\sum_{i=1}^{n}\frac{(\pi_i\rho_i)^{-1/c}}{\rho_i}.$$

式（35）便是在说最优配置由货币上的投资额

$$x=-\frac{a}{b}\sum_{i=1}^{n}\frac{1}{\rho_i}$$

以及共同基金上的投资额 W_0-x 构成。这个共同基金由份额为

$$\delta_i=\frac{(\pi_i\rho_i)^{-1/c}}{\rho_i}\Big/\sum_{i=1}^{n}\frac{(\pi_i\rho_i)^{-1/c}}{\rho_i}$$

的各种阿罗—德布鲁证券构成。

$$U'(W)=ae^{bW} \tag{ii}$$

通过和式（i）完全一样的代数运算，可以得到

$$W_i=rx'+\delta_i'\rho_i(W_0-x'), \quad i=1,\cdots,n \tag{35'}$$

其中

$$r=1\Big/\sum_{i=1}^{n}\frac{1}{\rho_i}, \quad x'=\frac{1}{b}\sum_{i=1}^{n}\frac{\log a\pi_i\rho_i}{\rho_i}+W_0$$

以及

$$\delta_i=\frac{\log a\rho_i\pi_i}{\rho_i}\Big/\sum_{i=1}^{n}\frac{\log a\rho_i\pi_i}{\rho_i}$$

同样，式（35′）可以被直接地解释为，最优配置可以通过购买货币和一种有风险共同基金来实现。

值得明确提出的是，式（35）和式（35′）的一个有趣的地方，即当共同基金按上面所述的方式形成时（显然不用这种特殊选择，也有许多其他方式去表示货币分离）。如果边际效用具有式（10）的形式，货币上的投资独立于初始财富，如果边际效用具有式（11）的形式时，风险共同基金上

的投资独立于初始财富。

五、没有货币时一般市场上的分离

我们转向分析，什么时候分离可以在这样的市场发生，在这个市场上证券不可用于覆盖所有的或有状态，也就是说证券的数量少于状态的数量，而且不存在货币，即不可能形成在所有状态下都有相同回报的共同基金。我们的方法仍旧是考虑所有这样的市场的一个子集，而且证明对于这个子集，只有一定类型的效用函数是可行的。然后，可以直接验证，这些类型的效用函数在没有货币的一般市场上也是可行的。[13]

为了进行分析，我们考虑特殊的市场子集，在这些市场上前 $n-1$ 种是阿罗—德布鲁证券，最后一种证券在剩下 $v-n+1$ 种状态中都有正的回报（令 v 为状态的个数，和前面一样，n 为证券的个数），从而回报的模式具有如下结构：

$$R=[\rho_{i\theta}]=\begin{bmatrix}\rho_1 & 0 & \cdots & 0 & 0 & \cdots & 0\\ 0 & \rho_2 & \cdots & & & & \\ \vdots & & & & & & \\ 0 & 0 & \cdots & \rho_{n,n} & \rho_{n,n+1} & \cdots & \rho_{n,v}\end{bmatrix} \tag{36}$$

其中 $2<n<v$，对某些 $\theta\neq\theta'$ 有 $\rho_{n,\theta}\neq\rho_{n,\theta'}$

特别注意到在式（36）刻画的市场上，既没有一种共同基金可以占优于任何初始证券，也没有一种共同基金具有货币的特征。因此通过分析这样的市场上允许卖空时的最优配置，我们能够证明：

定理 5.1：在一般市场上存在分离的充分必要条件是效用函数满足二次型式（10′）或者是常相对风险厌恶式（10″）。

定理 5.1 的证明：（必要性）　　由阿罗—德布鲁市场上分离的讨论可知，我们只需要考虑由式（8）和式（9）表示的效用函数类型。一种有效的方式是先考虑子情况式（10）和式（11），其次是一般的情形（除去子情况）式（8）和式（9）。我们的过程是验证并且看是否对于这些情况的每一种，在式（36）刻画的特殊市场上分离可以达到。对于我们的论证，$n=3$，

13　我们找到的效用函数在一般市场上可推出分离，当货币可用时，可推出货币分离，这一结果是人们已熟知的。在下一部分，我们将探究更深的问题，即是否存在别的效用函数（除了式（8）和式（9）列出的可能的效用函数之外），当货币可用时，能够推出分离或者货币分离。即货币在推出分离中本质上所起的作用。

$v=4$时的这种市场对应的情形已有足够的一般性。

$$U'(W) = (a + bW)^c \tag{i}$$

对于这一类特别的函数，条件式（3）可以被写作

$$(a + bZ_i\rho_i)^c \pi_i\rho_i = \lambda, \quad i = 1,2$$

$$\sum_{\theta=3}^{4} (a + bZ_3\rho_{3\theta})^c \pi_\theta\rho_{3\theta} = \lambda$$

或者，通过假定分离并代入需求函数如式（14）中形式，

$$(a + bA_i\rho_i W_0 + bB_i\rho_i h(W_0))^c \pi_i\rho_i = \lambda, i = 1,2$$

$$\sum_{\theta=3}^{4} (a + bA_3\rho_{3\theta}W_0 + bB_3\rho_{3\theta}h(W_0))^c \pi_\theta\rho_{3\theta} = \lambda \tag{37}$$

如果h（W_0）不是不变的常数，那么通过对式（37）的前两个方程关于W_0求两次导，就可以证明必然是如下情形：

$$\pi_1\rho_1 = \pi_2\rho_2$$

因此，不失一般性地，我们可以假定$h(W_0)$是不变的常数。但是，如果令$h(W_0)=K$，那么通过求解前两个方程中任何一个，都可以得到

$$\lambda = (\gamma + \delta W_0)^c \tag{38}$$

其中$\gamma=(a+bB_i\rho_iK)(\pi_i\rho_i)^{1/c}$且$\delta=bA_i\rho_i(\pi_i\rho_i)^{1/c}$，$i=1$，2。进一步地，当$h$（$W_0$）$=K$时，式（37）中的第三个方程就可以写作

$$\sum_{\theta=3}^{4} [a + bB_3\rho_{3\theta}K + bA_3\rho_{3\theta}W_0]^c \pi_\theta\rho_{3\theta} = \lambda$$

或者，令$\gamma_\theta=(a+bB_3\rho_{3\theta}K)(\pi_3\rho_{3\theta})^{1/c}$且$\delta_\theta=bA_3\rho_{3\theta}(\pi_\theta\rho_{3\theta})^{1/c}$，而且将式（38）代入，可有

$$\sum_{\theta=3}^{4} (\gamma_\theta + \delta_\theta W_0)^c = (\gamma + \delta W_0)^c \tag{39}$$

最终，通过定义$x=(\gamma+\delta W_0)^{-1}$，$\gamma'_\theta=\gamma_\theta-(\delta_\theta/\delta)\gamma$以及$\delta_\theta'=(\delta_\theta/\delta)$，式（39）可以被简化为

$$\sum_{\theta=3}^{4} (\delta_\theta' + \gamma_\theta' x)^c = 1 \tag{40}$$

在关于x求两次微分之后，可以得到

$$c(c-1)\sum_{\theta=3}^{4} (\delta_\theta' + \gamma_\theta' x)^{c-2} \gamma'^2_\theta = 0$$

因此（记得$U''<0$要求$c\neq0$，而且注意推导出式（40）要求$\delta_\theta'+\gamma'_\theta x>0$）或者有$c=1$，这种情况下效用函数是二次型式（10′），或者

$$\gamma_\theta{}' = 0,\text{也就是},\frac{\gamma_\theta}{\delta_\theta} = \frac{\gamma}{\delta},\text{也就是},\frac{a + bB_3\rho_{33}K}{bA_3\rho_{33}} = \frac{a + bB_3\rho_{34}K}{bA_3\rho_{34}} \tag{41}$$

但是，式（41）意味着或者是 $a=0$，在这种情况下，效用函数是常相对风险厌恶函数式（10′），或者 $\rho_{33}=\rho_{34}$，违背了我们对市场结构做的假定式(36)。

因此，我们已经证明了，给定式（36），式（10）要实现分离必须是式（10′）或者式（10″）这两种特殊情形。

$$U'(W) = ae^{bW} \tag{ii}$$

通过和前面实质上一样的论证过程，可以排除这类型效用函数的分离。特别地，重复上面论证的每一步，与式（40）不同的，我们得到了如下条件

$$\sum_{\theta=3}^{4} \delta_\theta{}' e^{\gamma} \theta'^{x} = 1 \tag{42}$$

其中 x，$\gamma_\theta{}'$和 $\delta_\theta{}'$是适当选取的变量和参数。我们省略这个论证的进一步的细节。

$$AU'(W)^\alpha + BU'(W)^\beta = W \tag{iii}$$

对于 $n=3$，$v=4$ 时式（36）的回报模式，一阶条件式（3）可以被一般地写作

$$\begin{aligned} &U'(Z_i\rho_i)\pi_i p_i = \lambda, i = 1,2 \\ &\sum_{\theta=3}^{4} U'(Z_3\rho_{3\theta})\pi_\theta\rho_{3\theta} = \lambda \end{aligned} \tag{43}$$

假定边际效用具有式（8）的形式而且需求函数具有式（14）的形式，很容易看到，由式（43）的前两个方程，我们可解出用 λ^α 和 λ^β 表示的 W_0 和 $h(W_0)$，比如说

$$W_0 = C_1\lambda^\alpha + C_2\lambda^\beta$$

$$h(W_0) = C_3\lambda^\alpha + C_4\lambda^\beta$$

那么式（43）的第三个方程可以表示为

$$U'(W_3)\pi_3\rho_{33} + U'\left(\frac{1}{\gamma}W_3\right)\pi_4\rho_{34} = \lambda \quad \text{同时} \quad W_3 = \delta\lambda^\alpha + \epsilon\lambda^\beta \tag{44}$$

其中 $\gamma=\rho_{33}/\rho_{34}$，$\delta=\rho_{33}$（$A_3C_1+B_3C_3$）且 $\epsilon=\rho_{33}$（$A_3C_2+B_3C_4$）。我们将会证明式（44）对于式（8）这一类型中任意给定的效用函数一般是不成立的(当 $A\neq0$，$B\neq0$ 且 $\alpha\neq\beta\neq0$ 时，即除了式（10）、式（10′）和式（10″）可能成立，但我们已独立考虑过)。

令 $u = U'(W_3), v = U'\left(\frac{1}{\gamma}W_3\right)$。结合式（8）和式（44）可以得到 u，v 和 λ 表示的系统方程

$$Au^{\alpha} + Bu^{\beta} = \delta\lambda^{\alpha} + \epsilon\lambda^{\beta}$$

$$Av^{\alpha} + Bv^{\beta} = \frac{1}{\gamma}[\delta\lambda^{\alpha} + \epsilon\lambda^{\beta}] \tag{45}$$

$$u\pi_3\rho_{33} + v\pi_4\rho_{34} = \lambda$$

它们必须在一定的非负的区域上成立。通过式（45）中前两个方程求解 $u^{\beta-\alpha}$ 并且对结果取等，我们可以得到

$$\frac{B}{A}u^{\beta-\alpha} = \frac{1-\gamma\left(\frac{v}{u}\right)^{\alpha}}{1-\gamma\left(\frac{v}{u}\right)^{\beta}} = \frac{1-\frac{\delta}{A}\left(\frac{\lambda}{u}\right)^{\alpha}}{1-\frac{\epsilon}{B}\left(\frac{\lambda}{u}\right)^{\beta}} \tag{46}$$

对 λ 进行适当的重新定义，式（45）中最后一个方程可以被改写为

$$\mu + (1-\mu)\frac{v}{u} = \frac{\lambda}{u} \tag{47}$$

令 $x = v/u$。最后把式（47）代入式（46）可以得到由 x 表示的方程

$$\frac{1-\gamma x^{\alpha}}{1-\gamma x^{\beta}} = \frac{1-\frac{\delta}{A}(\mu+(1-\mu)x)^{\alpha}}{1-\frac{\epsilon}{B}(\mu+(1-\mu)x)^{\beta}} \tag{48}$$

它必须在一定的非负区间成立（因为由式（46），x 不可能是常数，除非 u 是常数）。但是这意味着对于满足如下定义的解析函数 f

$$f(x) = (1-\gamma x^{\alpha})\left[1-\frac{\epsilon}{B}(\mu+(1-\mu)x)^{\beta}\right] - (1-\gamma x^{\beta})\left[1-\frac{\delta}{A}(\mu+(1-\mu)x)^{\alpha}\right] \quad \text{当 } x>0 \text{ 时，} \tag{49}$$

下式必须成立：

$$f(x) = 0 \quad \text{对所有 } x>0 \text{ 成立} \tag{50}$$

通过考虑 x 的特殊取值（即 $x=1$，$x=(1/\gamma)^{1/\alpha}$ 和 $x=(1/\gamma)^{1/\beta}$），容易证明式（50）一般不成立。[14]

因此，我们已经证明了式（8）在没有货币的一般市场上并不具有分

14　确实，只有当 $\gamma=1$ 或者 $\rho_{33}=\rho_{34}$ 时它才会成立，然而这已经被假定式（36）排除。我们顺便指出，正是在这一步，$\alpha=0$ 或 $\beta=0$ 以及 $\alpha=\beta$ 这两种情形的论证失败。

离性。

$$U'(W)^{\alpha}(A+B\log U'(W))=W \quad \text{(iv)}$$

再次，基本上重复前面的论证，我们可以排除这类型效用函数的分离，因此我们将其忽略。（充分性）这个论证仅由下面几步构成：写下对应每一个方程的预算约束式（1）和一阶条件式（3），然后求解 λ 和 Z_i 的 $n+1$ 个方程来确认式（14）。这个演算过程将会在第八部分中详细给出，在那里我们分析了具有分离性的不同效用函数对应的需求函数的精确形式。

六、存在货币的一般市场上的分离

在上一部分中，我们发现在完全一般的市场上，只有二次型或常相对风险厌恶效用函数能够推出分离性。我们考虑这样一个问题——它的重要性在于存在无风险资产是一个非常普遍的简化假设——货币的存在（或者初始的证券或者共同基金）对分离将有什么意义。它的影响可能有两种：首先，也许货币的存在允许其他类型的效用函数（显然除了式（8）和式（9）描述的那些之外）也能推出分离，而不管是不是货币分离。确实仔细阅读定理5.1的必要性的论证将会发现对于式（i）至式（iv）的每一类情形，证明的过程都和形式有关，比如（i），如果市场具有特殊结构（36），但是对 $\rho_{n\theta}$ 没有任何限制，如果效用函数属于式（10）这一类型且 $a\neq0$，$c\neq0$，要使分离达到，必然要求 $\rho_{n\theta}=\rho_{n\theta'}$，对于所有 $n\leqslant\theta\leqslant\theta'\leqslant v$ 都成立。也就是说，前面部分证明了或者我们不对第 n 种证券施加限制，那么式（36）刻画的市场必须退化到在这个市场上货币是可用的，[15]或者分离无法达到。因此，我们不能确认是否货币的存在将允许式（10′）和式（10″）之外的函数具有分离性。当然还有第二种影响：货币的存在可能导致货币分离。

我们证明的结论是，货币的存在只是略微扩展了具有分离性的函数类型——包含了函数类型如式（10），其中二次型和常相对风险厌恶函数是它的特例，以及常绝对风险厌恶函数式（11）——而且任何时候如果货币存在且分离达到，货币分离都将达到。也就是，很粗略地说，货币的存在产生的唯一重要区别是它确保了任何时候如果分离存在必有货币分离存在。我们将这一基本结果正式表述于下面的定理：

15 显然，这一点上有含糊之处，我们也可以这样解释这个结果，即，说成是“或者不对第 n 种证券施加限制时，式（36）刻画的市场必须退化到阿罗—德布鲁市场，或者……”正文的解释是最不利于我们当前讨论的。

定理6.1：在存在货币的一般市场上，货币分离的充分条件，同时也是分离（也就是货币分离）的必要条件是效用函数满足式（10）或式（11）。

定理6.1的证明：正如前面提到的，充分性是人们已熟知的结果，需要证明的是必要性。但是这个证明过程基本上是上一部分定理5.1的证明过程的一个更细致的版本。这里，我们把兴趣限制在一个特别的市场，它的回报模式具有式（36）的结构同时加上了一种安全资产：[16]

$$R = [\rho_{i\theta}] = \begin{bmatrix} \rho_1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \rho_2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & & \\ 0 & 0 & 0 & \cdots & \rho_{n-1,n-1} & \rho_{n-1,n} & \cdots & \rho_{n-1,v} \\ r & r & r & \cdots & r & r & \cdots & r \end{bmatrix}$$

$$\text{其中对某些 } 3 < n < v, \rho_{n-1,\theta} \neq \rho_{n-1,\theta'}, \theta \neq \theta' \qquad (36')$$

然后证明对于这个市场子集（对于存在货币的任意市场更是如此），要达到分离（也就是货币分离）效用函数不可能是式（8）或式（9）（除非它退化到式（10）或式（11））。事实上，这两种情形的论证和以前非常相像，然而，麻烦也很多，我们将要做的所有事情，只是概述和前面论证必然不同的差异。[17]

$$AU'(W)^\alpha + BU'(W)^\beta = W \qquad (iii')$$

此时不存在 W_3 和 W_4 的简单联系（以前有 $W_3 = \gamma W_4$）。它导致式（45）中方程组转化为更一般的方程组

$$\begin{aligned} Au^\alpha + Bu^\beta &= \delta_u \lambda^\alpha + \epsilon_u \lambda^\beta \\ Av^\alpha + Bv^\beta &= \delta_v \lambda^\alpha + \epsilon_v \lambda^\beta \\ \xi u + nv &= \lambda \end{aligned} \qquad (45')$$

对于适当定义的变量和参数成立，然而通过一个直接但很繁杂的证明可以推出式（45′）实际上等价于式（45）。[18]

---

16 注意到我们还是让 $n < v$，即，我们并不在阿罗—德布鲁市场上。

17 为了保证当前的论证有足够的一般性，市场表达式（36′）的最小的维度显然是 $n = 4$，$v = 5$。

18 这个证明本质上是把式（45′）简化为下面这个类似式（49）的式子，然后证明如果式（50）成立，$W_3 = \gamma W_4$ 也将成立。

$$f(x) = (\delta_u/A - x^\alpha)[\epsilon_v/B - (\mu + (1-\mu)x)^\beta] - (\epsilon_u/B - x^\beta)[\delta_v/A - (\mu + (1-\mu)x)^\alpha] \qquad (49')$$

$$U'(W)^{\alpha}(A + B\log U'(W)) = W \quad \text{(iv')}$$

这在本质上和情形式（iii'）一样。

我们应该强调的是，我们试图把在没有货币的市场上得到式（10）和式（11）不能推出分离的证明用在存在货币的市场上，期望能得到类似结论，但这种努力是失败的。这当然是定理的一个中心部分。也许值得进一步指出的是，这种努力失败在最后一步，以式（10）的论证为例，就是想要类似于式（41）那样推出矛盾 $\rho_{33} = \rho_{34} = \rho_{35}$。

七、广义分离

到目前为止，我们考虑的效用函数都是使分离对应两种共同基金，而且我们发现，粗略地说，只有很少数效用函数具有这样的分离性。那么，人们很自然要问，仅有两种共同基金的要求限制性有多强。下面的定理是关于阿罗—德布鲁市场上广义分离[19]的可能性，它强烈支持这样一个猜测：分离性最有限制性的特征不是对共同基金数量的限制，而是存在共同基金这一要求。这个定理也很清楚地说明了为什么我们没有对广义分离（与分离相对比）作进一步的分析。

定理 7.1：给定阿罗—德布鲁证券，广义分离的充要条件是边际效用的反函数 $G \equiv U'^{-1}$ 可以被表示为

$$G(xy) = \sum_{i=1}^{m}\sum_{j=1}^{m} f_i(x) C_{ij} f_j(y) \quad (51)$$

其中

$$f_i(x) = \sum_{k=1}^{K} x^{-\alpha_k} \sum_{l=1}^{L_k} D_{ikl} \log^{l-1} x \text{ 同时 } 2 \leqslant m < n, 1 \leqslant L_k \text{ 而且 } \sum_{k=1}^{K} L_k = m$$

注释：这个定理的含义是，甚至在阿罗—德布鲁市场上，广义分离也是局限于非常少数的效用函数的性质。特别要注意的是式（51）的暗含条件是参数 C_{ij} 和 D_{ikl} 必须满足一定关系使得当 xy = 常数时能推出 G（xy）= 常数。例如，很容易证明对于 $m = 3$ 时，也就是当所有相关的市场机会可以被最多三种共同基金提供时，边际效用必须满足

$$AU'(W)^{\alpha} + BU'(W)^{\beta} + CU'(W)^{\gamma} = W$$
$$U'(W)^{\alpha}(A + B\log U'(W)) + CU'(W)^{\gamma} = W$$

---

19 回顾我们早先的定义，也就是指这种情形，即一定的效用函数使得所有的市场机会可被至多 $2 \leqslant m < n$ 种共同基金提供。

或者

$$U'(W)^{\alpha}(A + B\log U'(W) + C\log^2 U'(W)) = W$$

定理 7.1 的证明：这个证明只是第三节和这篇论文的附录的论证的推广。除了最后的少数几步外，[20] 如果不说细节，那么在精神上它和那些论证也是一样的——也就是说，所有在推广的情形中本质上有趣的东西都已包含在前面给出的论证中——所以我们不再赘述。

八、分离的公式

前面论证的大部分致力于说明，只有非常有限的几类函数式（10′）和式（10″）或者（假定货币存在）式（10）和式（11）在完全一般的证券市场上能够具有分离性。完整起见，我们这里简要列出精确公式，它们刻画了当投资者具有这些类型中某一种效用函数代表的偏好时会如何行为。

定理 8.1：给定式（10′）或式（10″），或者货币存在时式（10）或式（11）成立，对于每一种证券的需求是关于初始财富呈线性的

$$Z_i = A_i W_0 + B_i, \quad i = 1, \cdots, n \tag{52}$$

对于式（10′），参数 A_i 和 B_i 由下式给出

$$\begin{aligned} A_i &= -\sum_{j=1}^{n-1} \sigma_{ij}^{-1} E\rho_{n\theta}(\rho_{i\theta} - \rho_{n\theta}), \quad i = 1, \cdots, n-1 \\ &= 1 - \sum_{j=1}^{n-1} A_j, \quad i = n \end{aligned} \tag{53}$$

$$\begin{aligned} B_i &= -\frac{a}{b}\sum_{j=1}^{n-1} \sigma_{ij}^{-1} E(\rho_{i\theta} - \rho_{n\theta}), \quad i = 1, \cdots, n-1 \\ &= -\sum_{j=1}^{n-1} B_j, \quad i = n \end{aligned}$$

其中 σ_{ij}^{-1} 是元素为 $E(\rho_{i\theta} - \rho_{n\theta})(\rho_{j\theta} - \rho_{n\theta})$ 的矩阵的逆矩阵的代表元素，对于式（10″），可有

A_i，$i = 1$，…，n 是下式的解

20　广义问题的结构，要求把类似于式（Ⅱ.15）中右边的矩阵 $\begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix}$ 重新写成约当（Jordan）标准型——而不是对式（Ⅱ.15）两边都求微分——然后求解微分方程的生成系统。

$$\begin{cases} E(\rho_{n\theta} + \sum_{i=1}^{n-1} A_i(\rho_{i\theta} - \rho_{n\theta})^c(\rho_{j\theta} - \rho_{n\theta}) = 0, j = 1, \cdots, n-1 \\ A_n = 1 - \sum_{i=1}^{n-1} A_i \end{cases} \tag{54}$$

$$B_i = 0, i = 1, \cdots, n$$

对于式（10），可有

$$A_i = C_i br, \quad i = 1, \cdots, n-1$$

$$= 1 - (\sum_{i=1}^{n-1} C_i) br, \quad i = n$$

$$B_i = C_i a, \quad i = 1, \cdots, n-1$$

$$= -(\sum_{i=1}^{n-1} C_i) a, \quad i = n$$

其中

$C_i, i = 1, \cdots, n-1$ 是下式的解

$$E[1 + b] \sum_{i=1}^{n-1} C_i(\rho_{i\theta} - r)^c(\rho_{j\theta} - r) = 0, j = 1, \cdots, n-1 \tag{55}$$

而对于式（11），可有[21]

$$A_i = 0, \quad i = 1, \cdots, n-1$$

$$= 1, \quad i = n$$

$B_i, i = 1, \cdots, n$ 是下式的解

$$\begin{cases} E\exp(b \sum_{i=1}^{n-1} B_i(\rho_{i\theta} - r))(\rho_{j\theta} - r) = 0, \quad j = 1, \cdots, n-1 \\ B_n = -\sum_{i=1}^{n-1} B_i \end{cases} \tag{56}$$

定理 8.1 的证明：只要直接代入一阶条件式（3）的适当形式便可验证这些公式的正确性。例如，给定式（10′）这一类型的效用函数，一阶条件为

$$E(a + b \sum_{i=1}^{n} Z_i \rho_{i\theta}) \rho_{j\theta} = \lambda, \quad j = 1, \cdots, n \tag{57}$$

21　注意到第 n 种证券在式（53）和式（54）中是任意的，但是在式（55）和式（56）中必须是货币（它的回报仍旧表示为 r）。然而，在这四种情形中当安全资产存在时可以选货币为第 n 种证券。

利用约束条件式（1）（以去掉 Z_n）和式（57）中的最后一个方程（以去掉 λ），式（57）的前 $n-1$ 可以被写作

$$E(a + b\rho_{n\theta}W_0 + b\sum_{i=1}^{n-1} Z_i(\rho_{i\theta} - \rho_{n\theta}))(\rho_{j\theta} - \rho_{n\theta}) = 0, \quad j = 1,\cdots,n-1 \tag{58}$$

很容易看出式（52）和式（53）的定义给出的 Z_i 满足式（1）和式（58），所以也就满足式（1）和式（3）。

其他三种情形是完全类似的，它们的推导留给感兴趣的读者去做。

还有别的一些东西值得评论。首先，利用式（52）和式（53），很容易看出对于任何二次型效用函数，共同基金可以被选定为分别使回报的方差和变异系数最小。[22]其次，从式（52）和式（54），可以直接证明对于任何给定的常相对风险厌恶函数只需要一种共同基金，即证券按照 A_i 的比例生成的基金。再次，对于任何给定的常绝对风险厌恶函数，任何特定的风险证券上的投资额是常数，而且它占所有风险证券上的总投资额的比例独立于效用函数的参数。

最后，我们指出定理 8.1 结合分离的一般表述式（14）可以直接推出

定理 8.2：(i) 在一般市场（存在货币的市场）上分离存在的充分必要条件是任何证券的需求函数都是关于财富是线性的；(ii) 任何证券的需求函数都和初始财富成比例的充分必要条件是效用函数是常相对风险厌恶式（10″）；任何证券的需求函数关于初始财富都是线性的充分必要条件是效用函数是二次型式（10′）（或者满足式（10）或式（11））。

九、回报的概率分布类型

一个效用函数为二次型的投资者的投资行为，事实上就像他在评估他的投资组合的回报的均值和方差。[23]这意味着对于这种特殊情形，分离可被解释为只是反映了这样一个事实：所有的有效的投资组合（即给定回报均值时方差最小）都可以被一对共同基金生成。

正如托宾在他的开创性文章中指出的那样，要达到对投资组合的配置的均值—方差分析，我们可以假定一个正态分布的回报和任意给定的效用函

22　这只反映了这样的事实，即具有二次型效用函数的投资者真正感兴趣的只是他的投资组合回报的均值和方差；可参见下一节。

23　正如我们一直假定的，当投资者关心总回报（而不仅是回报率）时，这种评估显然依赖初始财富。

数，而不是一个二次型的效用函数和任意模式的回报。这自然在暗示这样一个问题，即是否存在别的类型的分布也可以得出均值—方差分析从而实现投资组合的分离？

通过考虑下面的简单例子容易看出它并不成立，令 $n=2$，且 $\rho_{i\theta}$，$i=1$，2，是独立同分布的随机变量。要使均值—方差分析在这里可行，必须要求任何 $\rho_{1\theta}$ 和 $\rho_{2\theta}$ 的线性组合只是在比例和位置的参数上和 $\rho_{1\theta}$（或 $\rho_{2\theta}$）的初始分布有区别。但是使它能成立的分布类型只有稳定分布或帕累托—利维（Pareto-Levy）分布，而唯一具有有限方差的成员是正态分布。

进一步，可以直接证明如果（i）每一种风险证券的回报都是有相同特征指数的稳定分布；[24]（ii）而且存在货币，那么有效投资组合（这里是给定回报均值时有最小离差）可以被货币和另一种共同基金生成，当没有货币时这样的分离只能在正态分布时得到。

因此，即使将“风险”的概念扩展也不会对我们的结论有丝毫影响，即，在本质上，只有回报是正态分布时才会得到投资组合的分离。

十、结论

在本文中，我们已经证明了以下的论断是等价的。

一般而言：

A.（i）效用函数是常相对风险厌恶的。

（ii）所有相关的市场机会可以被唯一一种共同基金提供（即一个证券的固定组合）。

（iii）对任何一种证券的需求都和初始财富成比例。

B.（i）效用函数是二次型或常绝对风险厌恶的。

（ii）所有相关的市场机会可以被一对基金提供。

（iii）对任何一种证券的需求都是关于初始财富线性的。

如果存在货币（即一种在所有或有状态都产生相同回报的资产）：

C.（i）效用函数满足式（10）或式（11）。

（ii）所有相关的市场机会都可以被一对共同基金提供，其中一种被选定为货币。

（iii）对任何证券的需求都是关于初始财富线性的。

---

24 注意到要使回报均值有限，后者必然超过 1，然而这会导致本不希望有的推论（给定有限责任的机构），即回报可以是任何数（而不仅是非负的数）。

D.（i）效用函数满足

$$AU'(W)^{\alpha} + BU'(W)^{\beta} = W \quad \text{或者} \quad U'(W)^{\alpha}[A + B\log U'(W)] = W$$

（ii）所有相关的市场机会都可以被一对共同基金提供。

值得强调的是，我们的研究得到两个非常一般的结论。首先，能够推出分离的条件确实非常的苛刻（包含了在分析投资组合配置以及更一般地分析面对风险的行为时常常用到的效用函数）。其次，“分离”的含义可能会让人误解；当货币的存在提高了这种可能性，即用两种共同基金来代表和最优投资组合相关的所有市场机会——而且因此能够使配置风险资产的决策与配置安全资产和风险资产的决策分离——但这种提高是微不足道的。

附录Ⅰ

在本附录中，我们分析任意的证券集和阿罗—德布鲁证券之间的联系。这种联系的关键是证券的数量相对于自然状态的数量。为了能够比较，假定自然状态的数量是有限的，[25]同时令状态的数量为 v，此时存在三种可能性：（i）$v<n$；（ii）$v=n$；（iii）$v>n$。考虑第一种情况，$v<n$。在这种情况下，最多可以有 v 种证券的回报是线性独立的。这意味着至少有 $n-v$ 种证券的回报是被占优于或等价于其他证券组成的某种共同基金的回报。为了看清这一点，假定一个证券的子集 $i=1, \cdots, n'(\leqslant v+1)$，其中权重 $\alpha_i \neq 0$ 满足

$$\sum_{i=1}^{n'} \alpha_i \rho_{i\theta} = 0 \tag{Ⅰ.1}$$

那么或者 $\sum_{i=1}^{n'} \alpha_i > 0$，或者 $\sum_{i=1}^{n'} \alpha_i = 0$（因为我们只要求权重之间的比例是唯一的）。在这两种情况下，我们可以选择 $\alpha_1 > 0$ 并写成：

$$\rho_{1\theta} = \sum_{i=2}^{n'} -\frac{\alpha_i}{\alpha_1}\rho_{i\theta} \text{ 其中} \sum_{i=2}^{n'} -\frac{\alpha_i}{\alpha_1} = 1 - \frac{\sum_{i=1}^{n'} \alpha_i}{\alpha_1} \leqslant 1 \tag{Ⅰ.2}$$

因此，如果 $\sum_{i=1}^{n'} \alpha_i = 0$，我们可以形成一种共同基金，使得它的回报与第一种证券期望的回报相等，如果 $\sum_{i=1}^{n'} \alpha_i > 0$，我们可以形成一种共同基金，使得它的回报占优于第一种证券期望的回报，这可以简单地通过在式（5）中选择权重如下

25　在这个框架下，连续统的状态的结论和有限但是大量的状态的结论是一样的。

$$\delta_i = -\frac{\alpha_i}{\alpha_1}, \quad i = 2, \cdots, n' - 1$$

$$= 1 + \sum_{i=2}^{n'-1} \frac{\alpha_i}{\alpha_1} \quad i = n' \tag{I.3}$$

$$= 0 \quad 其他$$

因此，我们的结论是 $v < n$ 的情形或者与允许卖空不一致或者导致最优投资组合配置中不必要的不确定性。更广泛地说，这个论证证明了卖空的假定几乎要求证券的回报是线性独立的；因此，方便起见，我们随后都假设回报的线性独立。[26]

第二种情况 $v = n$ 相对稍微复杂一些。它提出了这样的可能性，就是从投资者的角度上来看，初始证券可能是和阿罗—德布鲁证券等价。也就是说，因为证券和状态的数量相等，就有可能形成具有如下性质 n 种共同基金，即第 i 种共同基金的回报为如果状态 $\theta = i$ 发生 $\mu_{i\theta} = \rho_i > 0$，在其他情况下，$\mu_{i\theta} = 0$。为了确认这种可能性，我们首先观察到如果这样的基金可以形成，那么由于第 i 种证券的每一美元投入的回报的倒数（$1/\rho_i$）代表了状态 i 时的每一元财富的价格，则在这样的价格下，每一种初始证券的回报的价值必然也是一美元：

$$1 = \sum_{\theta=1}^{v} \frac{1}{\rho_\theta} \rho_{i\theta}, \quad i = 1, \cdots, n$$

或者，用矩阵表示，

$$e = RM^{-1}e \tag{I.4}$$

这里

$$e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, R = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1v} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2v} \\ \vdots & & & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nv} \end{bmatrix} \quad 以及 \quad M = \begin{bmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & & \rho_n \end{bmatrix}$$

给定初始证券的回报是线性独立的，那么便有（$1/\rho_i$）>0，$i = 1, \cdots, n$ 要成立，当且仅当

$$M^{-1}e = R^{-1}e > 0 \tag{I.5}$$

容易证明，这种对 R 进一步的限制：(i) 等价于存在唯一一组 n 个共同基

---

26 严格地说，允许卖空的假设并不要求我们排除线性相关，它只是允许形成等价的共同基金。但是，如果排除这种可能性我们的论证并不会失去一般性；投资者允许卖空时对一种证券和与之等价的共同基金是无差异的（假定不存在交易成本）。

金，且它们是阿罗—德布鲁证券：（ii）意味着存在唯一一个的共同基金，而且它是货币。

把定义作为阿罗—德布鲁证券的共同基金的权重表示为：

$$D' = \begin{bmatrix} \delta_{11} & \delta_{21} & \cdots & \delta_{n1} \\ \delta_{12} & \delta_{22} & \cdots & \delta_{n2} \\ \vdots & & & \\ \delta_{1n} & \delta_{2n} & \cdots & \delta_{nn} \end{bmatrix}$$

而把定义作为货币的共同基金的权重表示为：

$$d' = [\delta_1 \delta_2 \cdots \delta_n]$$

而且货币自身表示为$\mu_\theta = r > 0$，其中$\theta = 1, \cdots, v$。那么，这些命题可以被重新正式表述为：

命题Ⅰ.1：存在唯一的D，使得$D'R = M$，其中当$i = j$时，$m_{ij} > 0$，当$i \neq j$时，$m_{ij} = 0$，并且$D'e = e$成立，当且仅当$R^{-1}e > 0$。

命题Ⅰ.2：假如$R^{-1}e > 0$，则存在以唯一的d和$r > 0$，使得$d'R = re' > 0$以及$d'e = 1$。

命题Ⅰ.1的证明：（必要性）$D'R = M$，其中M的元素m_{ij}满足当$i = j$时，$m_{ij} > 0$，当$i \neq j$时，$m_{ij} = 0$，那么可以推出M^{-1}存在，而且M^{-1}的元素m^{ij}满足当$i = j$时，$m^{ij} > 0$，当$i \neq j$时，$m^{ij} = 0$，并且有$M^{-1}D' = R^{-1}$，结合$D'e = e$可以推出$M^{-1}e = M^{-1}D'e = R^{-1}e > 0$。（充分性）$R^{-1}e > 0$意味着存在唯一的$M$，其中$M$的元素$m_{ij}$满足当$i = j$时，$m_{ij} > 0$，当$i \neq j$时，$m_{ij} = 0$，使得$M^{-1}e = R^{-1}e$，这可推出存在唯一的$D'$，使得$D'R = M$存在，进而推出$D'e = MR^{-1}e = MM^{-1}e = e$。

命题Ⅰ.2的证明：$R^{-1}e > 0$说明$r = (1/e'R^{-1}e) > 0$，$d' = re'R^{-1}$有良好的定义，而且$d'R = re'R^{-1}R = re' > 0$，$d'e = re'R^{-1} = 1$。此外，$d$和$r > 0$的存在使得$d'R = re'$和$d'e = 1$要求$d' = re'R^{-1}$以及$1/r = e'R^{-1}e$。

我们清楚地说明，当$n = v$（同时条件（Ⅰ.5）满足）时，货币的回报，或者更精确地说，作为货币的共同基金由下面的公式给出：

$$r = \frac{1}{e'R^{-1}e} = \frac{1}{e'M^{-1}e} \qquad (\text{Ⅰ}.6)$$

前面的分析告诉我们对于$v = n$的情形，如果初始证券满足式（Ⅰ.5），那么它们等价于阿罗—德布鲁证券以及另一组，比如说$n-1$种初始证券加上货币组成的证券集。剩下的唯一的问题是，条件（Ⅰ.5）不满足将意味着

什么？答案很清楚；因为，粗略地说，$R^{-1}e$ 中的非正元素反映了如下事实，即财富在某些状态下，最差是一个免费商品（而最好是一个负价格的商品），这样一种情形必然意味着至少有一种证券被占优。因此，条件（Ⅰ.5）自然来自于我们允许卖空的假定。

这些结果和它们的解释对于 $v = n = 2$ 的情形，如图 3 所示。

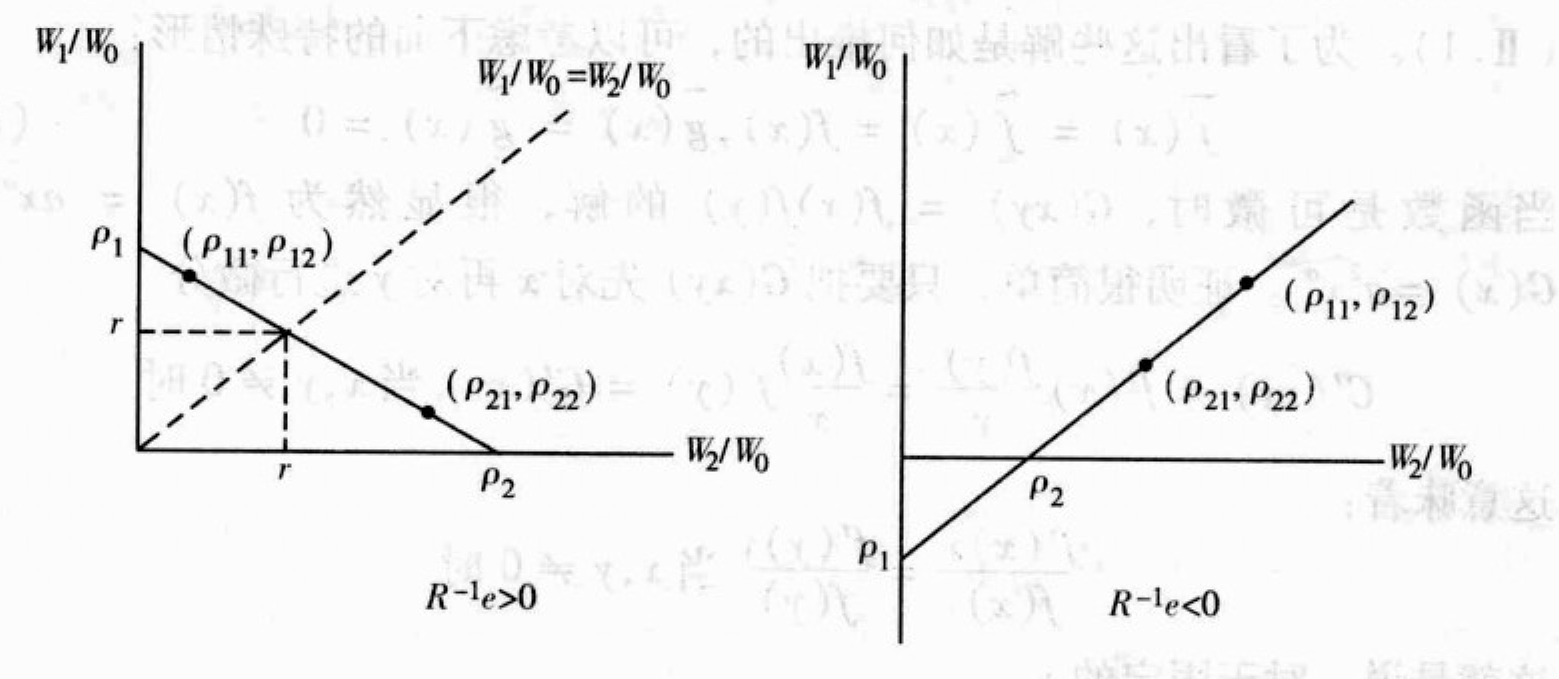


图 3

最后一种情形，$v > n$ 并没有太多可以说的，除了可能要强调这样一个事实，即当证券的数量少于状态的数量时，不可能为每一种或有状态提供保障。要一般性地描述投资组合行为，这种情形在某种意义上是最有趣的，也是对于我们想要分析投资组合分离这一特殊目的而言是最重要的。特别是，我们在正文中已经证明了，不可能为每一种或有状态提供保障将对具有分离性的效用函数类型加上非常苛刻的限制。

附录Ⅱ

在这个附录中，我们要证明正文第三部分中的命题，即函数方程：

$$G(xy) = \bar{f}(x)\tilde{f}(y) + \bar{g}(x)\tilde{g}(y) \qquad (Ⅱ.1)$$

的唯一解，当函数 G，$\tilde{f}$，$\tilde{g}$（$\bar{f}$，$\bar{g}$）在某些区间（$\underline{x}$，$\bar{x}$），$0 \leqslant \underline{x} \leqslant \bar{x} \leqslant \infty$ 上有定义且可微时，可以由下式给出：

$$\bar{f}(x) = \frac{\tilde{f}(x)}{A} = x^{\alpha}, \bar{g}(x) = \frac{\tilde{g}(x)}{B} = x^{\beta}$$

和

$$G(x) = Ax^{\alpha} + Bx^{\beta} \qquad (Ⅱ.2)$$

以及

$$\bar{f}(x) = \tilde{g}(x) = x^{\alpha}\left(\frac{A}{2} + B\log x\right)$$

$$\tilde{f}(x) = \bar{g}(x) = x^{\alpha}, G(x) = x^{\alpha}(A + B\log x)^{27} \qquad (\text{Ⅱ}.3)$$

在证明的开始，我们注意到式（Ⅱ.2）和式（Ⅱ.3）明显满足式（Ⅱ.1）。为了看出这些解是如何推出的，可以考虑下面的特殊情形：

$$\bar{f}(x) = \tilde{f}(x) = f(x), \bar{g}(x) = \tilde{g}(x) = 0 \qquad (\text{i})$$

当函数是可微时，$G(xy) = f(x)f(y)$ 的解，很显然为 $f(x) = ax^{\alpha}$，$G(x) = a^2x^{\alpha}$。证明很简单，只要把 $G(xy)$ 先对 x 再对 y 进行微分

$$G'(xy) = f'(x)\frac{f(y)}{y} = \frac{f(x)}{x}f'(y) = G'(xy) \text{ 当 } x, y \neq 0 \text{ 时}$$

这意味着：

$$\frac{f'(x)x}{f(x)} = \frac{f'(y)y}{f(y)} \text{ 当 } x, y \neq 0 \text{ 时}$$

这就是说，对于固定的 y，

$$\frac{d\log f(x)}{d\log x} = \alpha \neq 0, \text{当 } x \neq 0 \text{ 时}$$

这可以得到：

$$f(x) = ax^{\alpha}$$

$$\bar{f}(x) = \tilde{g}(x) = g(x), \tilde{f}(x) = \bar{g}(x) = 1 \qquad (\text{ii})$$

当函数可微时，$G(xy) = g(x) + g(y)$ 的解也很显然就是 $g(x) = a + b\log x$，

---

27 有几种方法可以写出（也是推出），式（Ⅱ.1）的解。例如，如果在式（Ⅱ.2）中，我们明确地把 A，B，α 和 β 写成复数，$A = A_1 + iA_2$，等等，我们便得到这样的表达式 $G(x) = x^{\alpha_1}(A_1\cos\alpha_2\log x - A_2\sin\alpha_2\log x) + x^{\beta_1}(B_1\cos\beta_2\log x - B_2\sin\beta_2\log x) + i\{x^{\alpha_1}(A_1\sin\alpha_2\log x - A_2\cos\alpha_2\log x) + x^{\beta_1}(B_1\sin\beta_2\log x - B_2\cos\beta_2\log x)\}$。但要使 $G(x)$ 为实数，参数必须满足：(a) $\alpha_1 = \beta_1$；而且 (b) 或者 $A_1 = -B_1$，$\alpha_2 = \beta_2$，或者 $A_1 = B_1$，$\alpha_2 = -\beta_2$；而且 (c) 或者 $A_2 = -B_2$，$\alpha_2 = \beta_2$，或者 $A_2 = -B_2$，$\alpha_2 = -\beta_2$。

而且当 $\alpha_2 = \beta_2$ 时，$G(x) \equiv 0$，而当 $\alpha_2 = -\beta_2$（也就是当 α，β 和 A，B 互为复共轭）时，

$$G(x) = 2x^{\alpha_1}(A_1\cos\alpha_2\log x - A_2\sin\alpha_2\log x)$$

再如，看上去是式（Ⅱ.2）的不同的表达式可以这样获得：用 $a+b$ 和 $a-b$ 来代替指数 α 和 β，$c+d$ 和 $c-d$ 代替相乘的常数 A 和 B，并用双曲函数代替指数函数 $x^b = c^{b\log x} = \sinh b\log x + c\cosh b\log x$，

$$G(x) = 2x^{\alpha}[d\sinh b\log x + c\cosh b\log x]$$

后来证明了，解这些特殊的表达式，是有帮助的。然而，基本上，在这个附录中我们试图进行这样的论证，它既避免不必要的参数表达，同时也阐明为什么额外的参数表达实际上是多余的。

$G(x)=2a+b\log x$。同样，把 $G(xy)$ 先对 x 再对 y 进行微分，我们可以得到：

$$G'(xy)=\frac{g'(x)}{y}=\frac{g'(y)}{x}=G'(xy)\text{ 当 }x,y\neq 0\text{ 时}$$

这可推出

$$g(x)=a+b\log x$$

$$\bar{f}=\frac{\tilde{f}}{A}=f\quad\text{其中}\quad\varphi(xy)=f(x)f(y)\qquad\text{(iii)}$$

$$\bar{g}=\frac{\tilde{g}}{B}=g\quad\text{其中}\quad\psi(xy)=g(x)g(y)$$

明显地，(a)$f(\varphi)$,$g(\psi)$ 的形式来源于（i），也就是说，$f(x)=ax^{\alpha}$（$\varphi(x)=a^2x^{\alpha}$）以及 $g(x)=bx^{\beta}$（$\psi(x)=b^2x^{\beta}$），然而（b）$A\varphi(xy)=f(x)f(y)$ 与 $B\psi(xy)=\bar{g}(x)\tilde{g}(y)$ 的和只是乘积 xy 的函数，就是说，$G(xy)=A\varphi(xy)+B\psi(xy)=Aa^2(xy)^{\alpha}+Bb^2(xy)^{\beta}$。由于在选择后一个方程的乘积常数时只有两个自由度，我们任意给定 A，B，令 a，b 等于1，就得到了式（Ⅱ.1）。

$$\tilde{f}=\bar{g}=f\quad\text{其中}\quad\varphi(xy)=f(x)f(y)\qquad\text{(iv)}$$

$$\bar{f}=\tilde{g}=fg\quad\text{其中}\quad\psi(xy)=g(x)+g(y)$$

此时 f 的形式必然可以从（i）推出，比如说 $f(x)=\beta x^{\alpha}$，然而 g 的形式必然可以从（ii）推出，比如说 $g(x)=(A/2)+B\log x$。此外，尽管现在 $\bar{f}(x)\tilde{f}(y)$ 与 $\bar{g}(x)\tilde{g}(y)$ 都不是 xy 的函数，它们的和，$G(xy)=\bar{f}(x)\tilde{f}(y)+\bar{g}(x)\tilde{g}(y)=\beta^2(xy)^{\alpha}(A+B\log xy)$，却是 xy 的函数。再次，因为在后一个方程中选择相乘的常数只有两个自由度，令 A 和 B 是任意的，同时 β 等于1，我们就得到了式（Ⅱ.2）。

得到式（Ⅱ.1）的解很容易；困难在于证明式（Ⅱ.2）和式（Ⅱ.3）是唯一的解。正如正文中式（Ⅱ.1）第一次出现时进行的论证，我们可以把这部分证明分为几个步骤。同时，方便起见，我们假定涉及的所有函数都有定义，而且对于自变量为正时是可微的；可以直接把论证过程进行必要的细化以包括这种情形，即 $0<U'(\infty)<U'(0)<\infty$。

（1）如果 $G(xy)=\bar{f}(x)\tilde{f}(y)+\bar{g}(x)\tilde{g}(y)$ 当 x，$y>0$，有一个解使得 $G\neq0$，那么或者（a）存在一个函数 f 使得

$$G(xy)=f(x)f(y)\qquad\text{(Ⅱ.4)}$$

或者（b）存在函数f和g，使得

$$G(xy)=f(x)g(y)+g(x)f(y) \tag{Ⅱ.5}$$

或者（c）存在函数f和g使得

$$G(xy)=f(x)f(y)+g(x)g(y) \tag{Ⅱ.6}$$

根据假设，我们有：对于所有x，$y>0$，

$$G(xy)=\bar{f}(x)\tilde{f}(y)+\bar{g}(x)\tilde{g}(y)=\bar{f}(y)\tilde{f}(x)+\bar{g}(y)\tilde{g}(x)=G(yx) \tag{Ⅱ.7}$$

对于任意y的两个取值，比如y_1，y_2，式（Ⅱ.7）可以得到方程

$$\begin{aligned}\tilde{f}(y_1)\bar{f}(x)+\tilde{g}(y_1)\bar{g}(x)&=\bar{f}(y_1)\tilde{f}(x)+\bar{g}(y_1)\tilde{g}(x)\\ \tilde{f}(y_2)\bar{f}(x)+\tilde{g}(y_2)\bar{g}(x)&=\bar{f}(y_2)\tilde{f}(x)+\bar{g}(y_2)\tilde{g}(x)\end{aligned} \tag{Ⅱ.8}$$

有两种情况需要考虑：

$$\tilde{f}(y_1)\tilde{g}(y_2)-\tilde{f}(y_2)\tilde{g}(y_1)=0 \text{ 对所有 } y_1,y_2>0 \text{ 成立} \tag{i}$$

对于这种情况，假如对所有$y>0$，$\tilde{f}(y)=0$，那么，由于$G\neq0$，必然存在至少一个值y_1，使得$\tilde{g}(y_1)\neq0$。因此，利用式（Ⅱ.8）的第一个方程，我们有：

$$\bar{g}(x)=\frac{\bar{g}(y_1)}{\tilde{g}(y_1)}\tilde{g}(x)\quad \text{对 } x>0 \text{ 成立}$$

把上式代入初始的函数方程式（Ⅱ.1），将得到式（Ⅱ.4）

$$G(xy)=\frac{\bar{g}(y_1)}{\tilde{g}(y_1)}\tilde{g}(x)\tilde{g}(y)=f(x)f(y)$$

其中，$f=(\bar{g}(y_1)/\tilde{g}(y_1))^{1/2}\tilde{g}$。另一方面，如果对某些$y_1>0$有$\tilde{f}(y_1)\neq0$，那么由假设$\tilde{f}(y_1)\tilde{g}(y_2)-\tilde{f}(y_2)\tilde{g}(y_1)=0$可推出

$$\tilde{g}(y)=\frac{\tilde{g}(y_1)}{\tilde{f}(y_1)}\tilde{f}(y)=C\tilde{f}(y),\text{对 } y>0 \text{ 成立}$$

再次，将上式代入式（Ⅱ.1）则产生式（Ⅱ.4）

$$G(xy)=(\bar{f}(x)+C\bar{g}(x))\tilde{f}(y)=h(x)\tilde{f}(y)=f'(x)f(y)$$

这里有$f=(h(y_1)/\tilde{f}(y_1))^{1/2}\tilde{f}$。

$$\tilde{f}(y_1)\tilde{g}(y_2)-\tilde{f}(y_2)\tilde{g}(y_1)\neq 0\text{,对某些 }y_1,y_2>0\text{ 成立} \tag{ii}$$

对于这种情况，我们通过式（Ⅱ.8）的一对方程把$\bar{f}$和$\bar{g}$求出，用$\tilde{f}$和$\tilde{g}$表示为

$$\begin{aligned}\bar{f}(x)&=C_1\tilde{f}(x)+C_2\tilde{g}(x)\\ \bar{g}(x)&=C_3\tilde{f}(x)+C_4\tilde{g}(x)\end{aligned} \tag{Ⅱ.9}$$

其中，

$$\begin{bmatrix}C_1 & C_2\\ C_3 & C_4\end{bmatrix}=\frac{1}{\tilde{f}(y_1)\tilde{g}(y_2)-\tilde{f}(y_2)\tilde{g}(y_1)}\times\begin{bmatrix}\bar{f}(y_1)\tilde{g}(y_2)-\bar{f}(y_2)\tilde{g}(y_1) & \bar{g}(y_1)\tilde{g}(y_2)-\bar{g}(y_2)\tilde{g}(y_1)\\ \bar{f}(y_2)\tilde{f}(y_1)-\bar{f}(y_1)\tilde{f}(y_2) & \bar{g}(y_2)\tilde{f}(y_1)-\bar{g}(y_1)\tilde{f}(y_2)\end{bmatrix}$$

注意到：(a) $G(y_1y_2)=\bar{f}(y_1)\tilde{f}(y_2)+\bar{g}(y_1)\tilde{g}(y_2)=\bar{f}(y_2)\tilde{f}(y_1)+\bar{g}(y_2)\tilde{g}(y_1)=G(y_2y_1)$ 成立，当且仅当 $\bar{g}(y_1)\tilde{g}(y_2)-\bar{g}(y_2)\tilde{g}(y_1)=\bar{f}(y_2)\tilde{f}(y_1)-\bar{f}(y_1)\tilde{f}(y_2)$，进而当且仅当 $C_2=C_3$；以及 (b)，$C_1=0$（$C_4=0$）要对所有 y_1，$y_2>0$ 成立，当且仅当 $\bar{f}(y_1)\tilde{g}(y_2)-\bar{f}(y_2)\tilde{g}(y_1)=0$（$\bar{g}(y_2)\tilde{f}(y_1)-\bar{g}(y_1)\tilde{f}(y_2)=0$）对所有 y_1，$y_2>0$ 成立。由于这最后的考虑，我们还得分析两种情况：

$$C_1=C_4=0 \tag{i}$$

在这种情况下，式（Ⅱ.9）变成

$$\begin{aligned}\bar{f}(x)&=C_2\tilde{g}(x)\\ \bar{g}(x)&=C_3\tilde{f}(x)\end{aligned} \tag{Ⅱ.10}$$

把式（Ⅱ.10）代入式（Ⅱ.1），利用条件 $C_2=C_3$，得到式（Ⅱ.5）

$$\begin{aligned}G(xy)&=C_2\tilde{g}(x)\tilde{f}(y)+C_3\tilde{f}(x)\tilde{g}(y)\\ &=f(x)g(y)+g(x)f(y)\end{aligned}$$

其中 $f=C_2\tilde{g}=C_3\tilde{g}$，$g=f$。

$$C_1\neq 0\quad\text{或者}\quad C_4\neq 0 \tag{ii}$$

不失一般性，我们可以假设 $C_1\neq 0$。然后把式（Ⅱ.9）代入式（Ⅱ.1），我们发现，这种情况将得到式（Ⅱ.6）

$$G(xy) = [C_1\tilde{f}(x) + C_2\tilde{g}(x)]\tilde{f}(y) + [C_3\tilde{f}(x) + C_4\tilde{g}(x)]\tilde{g}(y)$$

$$= C_1\tilde{f}(x)\tilde{f}(y) + C_2[\tilde{f}(x)\tilde{g}(y) + \tilde{f}(y)\tilde{g}(x)] + C_4\tilde{g}(x)\tilde{g}(y)$$

$$= C_1\left\{\left[\tilde{f}(x) + \frac{C_2}{C_1}\tilde{g}(x)\right]\left[\tilde{f}(y) + \frac{C_2}{C_1}\tilde{g}(y)\right]\right\} + \left[C_4 - \frac{C_2^2}{C_1}\right]\tilde{g}(x)\tilde{g}(y)$$

$$= f(x)f(y) + g(x)g(y)$$

其中$f = C_1^{1/2}$（$\tilde{f}$ +（C_2/C_1）$\tilde{g}$），$g =$（$C_4 -$（C_2^2/C_1））$^{1/2}\tilde{g}$。

我们已经刻画了式（Ⅱ.4）的解；这时我们应该转向分析式（Ⅱ.5）和式（Ⅱ.6）的解。进一步，因为式（Ⅱ.6）的分析几乎是式（Ⅱ.5）的分析等价的，我们只会详细说明后者的论证过程。

（2）在式（Ⅱ.5）的右边出现的函数f和g必须或者有如下的形式：

$$f(x) = ax^{\alpha} \quad 以及 \quad g(x) = bx^{\beta} \tag{Ⅱ.11}$$

其中G有式（Ⅱ.2）的形式；或者满足同样的微分方程：

$$Af(x) + 2Bf'(x)x + f''(x)x^2 = 0, 以及 Ag(x) + 2Bg'(x)x + g''(x)x^2 = 0 \tag{Ⅱ.12}$$

其中，G满足微分方程

$$AG(x) + BG'(x)x + G''(x)x^2 = 0 \tag{Ⅱ.13}$$

把式（Ⅱ.5）对x微商并把结果乘上x，我们得到：

$$G'(xy)xy = f'(x)xg(y) + g'(x)xf(y) \tag{Ⅱ.14}$$

式（Ⅱ.14）正好就是函数方程式（Ⅱ.1）（其中$G'xy = G$，$f'x = f$，等等）。因此，通过本质上是重复前一节的论证过程（因而我们省略掉细节），我们得出或者f和g的形式为式（Ⅱ.11），或者有存在$C_2' = C_3'$且$C_1' \neq 0$或者$C_4' \neq 0$

$$\begin{aligned} f'(x)x &= C_1'g(x) + C_2'f(x) \\ g'(x)x &= C_3'g(x) + C_4'f(x) \end{aligned} \tag{Ⅱ.15}$$

通过对这些方程微分，[28]我们有：

$$\begin{aligned} f''(x)x &= C_1'g'(x) + (C_2' - 1)f'(x) \\ g''(x)x &= (C_3' - 1)g'(x) + C_4'f'(x) \end{aligned} \tag{Ⅱ.16}$$

此时，求解式（Ⅱ.15），把f'用g和g'表示，g'用f和f'表示，并把结果代入式（Ⅱ.16），整理得到的表达式，我们得到：

28　注意到假如f和g一阶可微，那么式（Ⅱ.15）说明它们是解析的。

$$Af(x)+2Bf'(x)x+f''(x)x^2=0 \text{ 和 } Ag(x)+2Bg'(x)x+g''(x)x^2=0 \tag{Ⅱ.12}$$

其中 $A=C_2'C_3'-C_1'C_4'$，$2B=1-C_2'-C_3'$。最后，通过式（Ⅱ.14）关于 x 求微分，把结果乘上 x，则有

$$G''(xy)xy=(f''(x)x^2+f'(x)x)g(y)+(g''(x)x^2+g'(x)x)f(y) \tag{Ⅱ.17}$$

把 A · 式（Ⅱ.5），B · 式（Ⅱ.14）和式（Ⅱ.17）相加，很容易看出，式（Ⅱ.12）可推出式（Ⅱ.13）。

（3）满足函数方程式（Ⅱ.5）（而且，式（Ⅱ.6）更是这样）的微分方程式（Ⅱ.12）和式（Ⅱ.13）的唯一解，将得到式（Ⅱ.2）和式（Ⅱ.3）描述的形式。

这对我们要刻画式（Ⅱ.13）的解已经足够了。令 $z=(G'(x)x/G(x))$。那么

$$\frac{dz}{dx}x=\frac{G'(x)x}{G(x)}-\frac{G'(x)^2x^2}{G(x)^2}+\frac{G''(x)x^2}{G(x)}$$

或者

$$\frac{G''(x)x^2}{G(x)}=-z+z^2+\frac{dz}{dx}x \tag{Ⅱ.18}$$

将式（Ⅱ.18）代入式（Ⅱ.13）并简化，得到

$$A+Cz+z^2+\frac{dz}{dx}x=0$$

其中 $C=B-1$，或者

$$-\int\frac{dx}{x}=\int\frac{dz}{A+Cz+z^2} \tag{Ⅱ.19}$$

式（Ⅱ.19）的解析解（因此也是式（Ⅱ.13）的解）取决于 $q=C^2-4A$ 的符号：[29]

$$\begin{aligned}\log x &= -\frac{2}{\sqrt{-q}}\tan^{-1}\frac{2z+C}{\sqrt{-q}}+\text{积分常数},\text{当 } q<0 \text{ 时}\\ &= \frac{2}{2z+C}+\text{积分常数},\text{当 } q=0 \text{ 时}\\ &= -\frac{2}{\sqrt{q}}\tanh^{-1}\frac{2z+C}{\sqrt{q}}+\text{积分常数},\text{当 } q>0 \text{ 时}\end{aligned} \tag{Ⅱ.20}$$

29 实际上，式（Ⅱ.1）的解的形式对 q 的符号的依赖很像二次方程的解的形式对判别式的依赖。特别是，$q<0$ 对应式（Ⅱ.2）中的复共轭指数（也包括相乘的常数）；而 $q>0$ 对应式（Ⅱ.2）中实的，不相等的指数（也包括相乘的常数）。然而，$q=0$ 与式（Ⅱ.3）的联系是式（Ⅱ.2）的一个特殊情况，并不能做这样一种简单类比。

我们只需要考虑前两种情形，因为有关第三种情形的论证只是对于第一种情形论证的重复，仅把三角函数方程用双曲线方程（*hyper bolic*）

$$\log x = -\frac{2}{\sqrt{-q}}\tan^{-1}\frac{2z+C}{\sqrt{-q}} + \text{积分常数，当 } q < 0 \qquad \text{(i)}$$

来代替。

首先求反函数，使得 z 作为 x 的函数

$$\frac{d\log G(x)}{d\log x} = z = \alpha_1 - \alpha_2 \tan(\alpha_2 \log x + b_1)$$

然后积分，我们得到解

$$\log G(x) = \log b_2 + \alpha_1 \log x + \log\cos(\alpha_2 \log x + b_1)$$

或者

$$\begin{aligned} G(x) &= b_2 x^{\alpha_1}\cos(\alpha_2 \log x + b_1) \\ &= 2x^{\alpha^1}(A_1\cos\alpha_2\log x - A_2\sin\alpha_2\log x) \end{aligned} \qquad (\text{Ⅱ}.21)$$

其中 $\alpha_1 = -(C/2)$，$\alpha_2 = -\sqrt{(-q/2)}$，$A_i$ 实际上是任意的积分常数，然而 α 和 β 以及 A 和 B 是复共轭时，式（Ⅱ.21）恰好是式（Ⅱ.2）的特殊形式（参见此附录开始时的脚注）。

$$\log x = \frac{2}{2z+C} + \text{积分常数} \qquad \text{(ii)}$$

再次求反函数并积分，我们此时得到解

$$\begin{aligned} G(x) &= b_2 x^{\alpha}\log b_1 x \\ &= x^{\alpha}(A + B\log x) \end{aligned}$$

其中 $\alpha = -(C/2)$ 而且 A，B 实际上是任意的积分常数。因此，这种情况使得式（Ⅱ.3）成为式（Ⅱ.1）的解。

参考文献

1. K. J. ARROW, "Aspects of the Theory of Risk-Bearing," Helsinki, 1965.
2. D. CASS AND J. E. STIGLITZ, "The Structure of Preferences and Wealth Effects in Portfolio Allocation," mimeographed, Cowles Foundation for Research in Economics, Yale University.
3. N. H. HAKANSSON, "Risk Disposition and the Separation Property in Portfolio Selection," *Journal of Financial and Quantitative Analysis*, Dec. 1969, pp. 401–416.
4. J. PRATT, "Risk Aversion in the Small and the Large," *Econometrica*, January-April 1964, pp. 122–136.
5. J. TOBIN, "Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies*, February 1958, pp. 65–86.

不确定性下的厂商理论

关于公司财务纯理论的几个问题：破产与收购*

本文讨论了破产、收购和不一致的期望对公司财务政策的意义，我们认为，在合理的假设下，存在一个最优的债务—权益比率（debt - equity ratio）。以前的研究证明了：在一般的条件下，如果不存在破产的可能性，那么企业的财务政策对企业的价值没有影响，同时也不存在最优的债务—权益比率。在一定的限制性条件下，不存在破产这一假设可以被删除。我们证明了，当这些限制性条件未得到满足，并且确实存在一定的破产风险的时候，如果企业发行太多的债券，企业的价值就取决于它的债务—权益比率；企业真正的决策（如投资和技术选择）不能与它的财务决策相分离；企业真正的决策在生产方面也有可能是无效率的。最后，我们讨论了被收购可能对企业财务政策的意义。

一、绪论

本文要论证：在合理的假设下，存在一个最优的债务—权益比率。以前的研究[1]已经证明，在很一般的条件下，如果不存在破产的可能性，那么企

---

*　"Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Take-Overs", *The Bell Journal of Economics and Management Science*, Vol. 3, No. 2 (Autumn, 1972).

本文根据在 Hakone 会议上的演讲修改而成（1970 年 6 月 25 - 26 日）。作者要感谢大会的组织者 H. Uzawa，以及大会的其他参与者，作者感谢他们的有益评论和讨论。同时，作者要感谢 R. Merton 对本文的前一版本的评论。本文所做的研究得到了古根汉基金会（Guggenheim Foundation）、美国国家科学基金会以及福特基金会的资助。

1　参见（Modigliani and Miller）［参考文献 14］，以及 Stiglitz［参考文献 21、24］。

业的财务政策对企业的价值没有影响，同时也不存在最优的债务—权益比率。在一定的限制性条件下，不存在破产这一假设可以被删除。但是当这些限制性条件不能得到满足，同时企业存在一定的破产风险的时候，如果企业发行太多的债券，那会怎样？我们将证明：企业的估价将取决于它的债务—权益比率（第三部分）；因此，会有一个最优的债务—权益比率存在（第四部分）；企业真正的决策（例如，投资和技术选择）不能和它的财务决策相分离（两者必须同时决策），并且企业的真正决策在生产方面可能是无效率的。

与近期大多数的模型[2]相比，本文详细讨论的例子，除了考虑了破产之外，还有一个区别：它明确考虑了期望的差异化。在我们看来，期望一致的假设是不现实的，它的一些推论也是不可靠的，[3]而且它也不能解释资本市场上一些重要的现象。其中的一个现象——收购——将在第五部分详细讨论。

期望一致的假设被广泛采用可能是因为要在那样的模型中引入异质期望（heterogeneous expectations）并得到简单的结果，显然是很难的，[4]也是因为我们感觉用异质期望可以很容易地解释很多的东西，或是因为许多可以用异质期望解释的现象也可以用其他原因解释。例如，不同的个人会拥有不同的投资组合，这可以由异质期望解释，也可以由个人对风险的不同态度来解释。两者都难以直接观察到，因此很难确定两者的相对重要性。

较早的文献也趋向于关注期望的差异化，[5]近年来的文献关注对风险的不同态度。不同的模型关注不同的方面，并不是因为两者之一才是“真实的解释”，而是为了要了解它们如何起作用——因此最好把它们分开研究——例如，为了理解对待风险的不同态度对于投资组合配置的影响，我

2　特别是从夏普—林特纳（Sharpe-Lintner）模型衍生出的那些模型［参考文献13、15、18］。

3　期望一致时的均值方差模型不仅意味着财务政策的无关性，同时也意味着财务政策无关的原因是因为存在一个单一的共同基金（包括了所有的债券和股票）。就像在阿罗—德布鲁模型中（［参考文献2、4］），唯一不能通过分散投资消除的风险是与商业周期相关的风险。若两个企业合并成一个企业，则新企业的价值就是两个旧企业价值之和。对这些论点的扩展讨论，请参见［参考文献23］。

4　Lintner［参考文献12］。

5　请注意一个很有意思的现象，无论是凯恩斯在解释流动性偏好时，还是奈特（Knight）在解释零回报下企业家职权的连续性时，都对差异化期望给予了很大的关注，亦可参见Tobin［参考文献28］。

们需要假设每个人的期望都是一致的，或者是相反。在下文将要给出的例子中，我们假设每个人对待风险的态度一样。为简单起见，我们假设了风险中性。

二、破产以及财务决策和真实决策的独立性

在本部分中，我们试图给出破产的定义，并且说明破产为何对企业行为分析有重要的作用。

破产的定义 在一个两期模型中（企业在第一期进行投资，在第二期获得回报——因此也清算资产），破产很容易被定义为：企业的收入少于它对债券持有者的应付债务。然而，在一个运营着的企业中，收入可以少于企业的债务，企业也不至于破产，因为企业可以借更多。显然，如果企业权益的价值为零（在有限责任条件下，这不能为负），企业就破产了。类似地，假设企业现在没有破产，如果其未来收入流的价值少于其债务的价值，[6] 企业也会破产。这是我在多期模型中证明财务政策无关性所用的定义[7]（这一论证要求没有人相信在某日某种自然状态下权益的价值会为零）。该定义引出两个问题。第一，债务的价值应由市价来衡量，还是用到期价值来衡量？我们可以想像下述情形：市价超过了到期价值，企业在下一期有能力清偿到期的债务。市价衡量的债务价值是不是有可能等于或超过当期的企业价值，因而企业的权益价值为零？答案是否定的。因为在下一期，在所有的状态下，会有一定的剩余（residual）留给股东。本期的权益价值一定为正，因而，企业本期价值一定超过本期债券的价值（以市价衡量）。[8] 第二，尽管这是破产的充分条件。但这是必要条件吗？投资计划存在破产概率的必要条件是：如果企业现在不破产，一些人相信在将来的某一日，某一自然状态下，企业的价值会少于其债券的价值（像“无风险债券”那样估价），但在企业正式破产之日，企业权益的价值可能为正。其原因将通过下面的例子中进行说明。[9]

假设企业有唯一的资产——现金 975 000 美元。企业债券形式的负债为

6 若存在阿罗—德布鲁证券，我们可以很容易地计算出未来收入流的价值，但如果不存在这样的证券，那就可能很难确定其价值。

7 Stiglitz［参考文献 21］。

8 在［参考文献 21］中，我们证明了如果在任一日期任一自然状态下权益的价值都不为零，那么权益的价值加上债券的市场价值就独立于企业的财务政策。

9 我需要感谢罗伯特·默顿（Robert Merton），因为是他最早提出这一问题，而且给出了这个例子。

1 050 000 美元，并承诺从现在开始对每一单位债券每一期支付 1 美元。利率是 10%，故债券的现值是 953 637 美元，少于现金的价值。该企业只有一个项目，它有 0.25 的概率带来 1 150 000 美元的收入，有 0.75 的概率带来 960 000 美元的收入。因而企业有很大的概率不能清偿债务。然而，本期权益的价值是正的，如果市场是风险中性的，那么权益的价值是（100 000 美元 × 0.25）/1.1 = 22 727 美元。然而，债券持有者的期望回报不是 1 050 000美元，而是1 050 00 美元 ×0.25 +960 000 美元 ×0.75 =982 500 美元，其现值是（982 500 美元）/1.1 =893 181 美元，大大少于在本期兑现债券所能获得的价值（即 953 637 美元）。试想一个债券合约约定：如果债券持有者可以证明企业的政策很有可能导致在不久的将来不能履行债务，他们可以在本期宣布企业“破产”。在此情形下，上述企业可能在本期被宣布破产，哪怕破产后企业的权益价值仍为正（975 000 美元 -953 637 美元 = 21 363 美元）。

债券持有者可能不能使企业宣布破产，除非企业已经不能清偿债务，也就是说，尽管企业很可能在将来不能清偿债务，并且事实上，企业现今的资产价值少于债务价值，债券持有者不能在当前强迫企业破产。例如，我们对上述例子稍作改动，假设企业仅有 800 000 美元的现金，但投资机会仍然有 0.25 的概率可获得 1 050 000 美元的收入，有 0.75 的概率收入 960 000 美元。债务的现值（953 637 美元）大于现金的数量（企业的资本），可能有人会认为企业净值为负，即“破产”了。事实上，我们可以观察到，如果市场风险中立，企业的所有收入的净值为：

$$\frac{960\ 000 \times 0.75 + 1\ 050\ 000 \times 0.25}{1.1} = 915\ 908(\text{美元})$$

少于债务的现值。然而，只要债券持有者未能强迫企业破产，权益的价值就与前面的例子中一样（如果他们能强迫企业破产，显然企业权益的价值为零）。

这几个例子符合破产的条件：在某日某些自然状态下，企业不能清偿其债务，权益的价值为零。但是，这些例子也表明了我们的定义不能解决破产的时机问题（timing），即将来不能清偿债务的可能性是否使得债券持有者可以强迫企业现在破产。据我们的推测，这取决于债务合约是如何签订的。同长期债券相比，这一问题对于短期债券相对较轻。对于短期债券，每个人在每一期可以决定是否兑现债券。对于长期债券，个人购买债券后，可能会由企业管理层的变化，这种变化在债券持有者（或者可能也是整个市场）

看来可能导致破产的概率很高。然而，债券持有者什么都做不了。[10]

债券持有者和权益所有者之间明显的利益冲突加剧了这些问题。因为在破产时，企业承担有限责任，所以企业只考虑它不破产的状态下的回报。[11]例如，在上述两例中，上马该项目符合企业股东的利益。如果不上马该项目，则股东在第一个例子中获得 21 363 美元，在第二个例子中什么都得不到；若上马该项目，他们可获得现值为 22 727 美元的收入（假设风险中立）。在这两种情况下，如果企业在第一期清算，债券持有者可以获得更多。尽管在其中一个例子中，债券持有者获得的偿付额少于债权的现值（但这仍大于企业上马项目时期望回报的现值）。请注意，在第一个例子中，企业持有 975 000 美元的初始现金，投资的期望真实回报为负（-8%）；然而，如果股东要对此做出决策，该项目就会上马。显然，这可能导致生产的无效率。[12]

破产对企业财务政策的意义 破产对企业财务政策有重要作用，原因有以下三个方面。第一，不存在破产意味着企业对债券所需支付的名义利率独立于债务—权益的比例。若存在破产的可能，债券便成为风险资产，企业借得越多，名义利率[13]越高。我们没有理由假设对所有企业或个人来说，名义利率是债务—权益比率的同一函数；事实正与此相反。[14]一个完整的理论必须对给定企业名义利率的决定做出解释。我们将在下文中尝试进行解释。

请注意，债券的如下特性对最优债务—权益比率的决定有重要作用：若企业不破产，债券就是支付一定金额的承诺；若企业破产，债券是获得一定比例资产的承诺。债务人和债权人可能（一般都会）对一个风险债券的回报的均值和方差会有不同的看法。债务人通常会对偿债更有信心。在债务人看来，他们所支付金额的均值高于债务人相信所能拿到的金额的均值（而他们支付金额的方差则小于债务人相信自己所能得到的金额的方差）；债权人认为，破产概率越高，（随着债务—权益比率的升高）两个均值之间的差

10 当然，企业可以做的行为可能存在法律限制。

11 在债券发行之前，很明显企业会考虑各个自然状态下的回报，因为这会影响到它借款时的条件。然而，获得借款后，企业仅关注于它不破产的状态下的回报（当然，若企业需从资本市场持续地获得资源，那么它仍需考虑所有状态下的回报）。

12 如果存在有限的破产概率，企业的价值最大化并不等同于权益价值的最大化，显然，企业真正关注的是后者。

13 名义利率即企业不破产时在各种自然状态下所支付的利率。

14 正如莫迪格利安尼和米勒在［参考文献 14］中所完成的。在下文我们将看到，他们的结论（就算有破产，债务—权益比率也不影响企业价值）极端依赖此假设。

距就越大。换句话说，随着债务—权益比率的升高，债务人所要支付的名义利率比他所认为的“公平的”利率上升得更快。

破产对企业的财务决策也很重要，因为正如我们已证明的，[15]不存在破产时，个人的消费机会集不受债务—权益比率的影响；若存在破产（除非在特殊情形下），上述论断不再正确。

例如，我们可以考虑一个标准的两期模型，其中企业在第一期进行投资，在第二期获得回报。再考虑两种不同的财务政策——在一种政策下，企业发行数量为 B 的债券；在另一政策下，企业不发行债券（因而企业不可能无法清偿债务）。很显然，在后一种政策下，边际购买证券的个人所获之回报（如果企业的价值保持不变）可以与前一种政策下企业不破产的自然状态下完全一致。但如果企业在某些状态 θ' 下破产，在一种政策下，个人的回报是零；而在另一种政策下，他的回报（在偿付债务之后）是负的。换句话说，发行风险债券的企业其实发行了两种债券。当企业不破产时，它只是发行了市场还不存在的一种证券，该债券与其他任何企业的证券类似。通过提供了一种新的证券，破产可能改变社会的消费机会集。

这样，要使破产的可能性不影响企业的市场估价，需要以下两类条件：

（1）破产在实质上不能产生一种新的证券；

（2）破产所产生的新证券未造成任何改变。

条件（1）可以在下述条件下得到满足：[16]

（1a）如果已经存在的证券之数量与自然状态的数量一样（阿罗—德布鲁模型，这已在其他文献中得到讨论，并被证明是不现实的）。[17]

（1b）如果破产所产生的特殊证券也可以由个人产生。这一情形要求个人在边际上可以购买证券，[18]并且可以做出有限责任安排，借此（在我们的两期模型中）在企业利润低于到期应名义支付的借款时，股本不能被完全退回，但对于个人没有其他影响。

然而，需要清楚的是，这类有限责任的安排受到范围的限制，因此，破

15 ［参考文献21、24］。

16 如果存在大量的企业，且它们的回报是完全相关的（即用莫迪格利安尼和米勒的术语来说，这些企业属于相同的风险类（risk class），那么第三种（但不大可能的）情形就会出现。如果在相同风险类中，有一家企业的债务—权益比率与我们模型中的企业所考虑的债务—权益比率一样，那么模型中的企业选择债务—权益比率的决策就不会影响可供选择的证券集，哪怕这家企业可能破产。

17 ［参考文献22］和［参考文献23］。

18 即以证券作为抵押物借得资金。

产的确会产生一种新证券的假设仍然成立。[19]

破产所产生的新证券不带来任何改变的一般情形仅存在一种，即在个人的投资组合中，每个人购买的风险证券比例都一样。可以出现这种情形的条件在所谓的投资组合分离理论中给出，这已在其他文献中得到讨论。[20]该理论需要对效用函数和（或）可选证券之回报的分布做出非常约束性的限定。另外，它要求所有的个人对企业的回报有相同的预期。我们最反对的正是最后一个假设。

最后，破产所带来的交易成本和信息成本也会使得企业的市场估价依赖于破产概率。在接下来的讨论中，我们将忽略这些成本——尽管它们很重要——因为我们要证明，即使在莫迪格利安尼和米勒所描述的“理想”世界中（不存在交易成本），如果存在有限的破产概率，市场估价通常也会取决于债务—权益比率。

三、给定投资时的最优债务—权益比率

在本部分中，我们要通过简单的例子证明企业估价对其财务政策的依赖性。我们将在下一部分中推导出最优的财务政策。

假设有两组人，分别以上标 a 和 b 来表示。[21]经济中存在两个企业，[22]分别以下标 1 和 2 来表示。最初，a 组人拥有第一家企业的所有股份，[23]且两组人都认为第一家企业的利润是风险性的。类似地，两组人都认为第二家企业的利润是完全无风险的，并且它的生产技术是规模报酬不变的。这使得完全无风险资产的利率被固定在 r^*。[24]我们利用传统的两期模型：[25]第一期进行投资，第二期获得回报，然后企业“清算解散”。我们感兴趣的是财务决策，因此，我们假设真实决策（real decisions）（即投资水平）已经确定。

我们关注第一家企业。[26]在状态 θ 下，回报为 $X(\theta)$。对于任一给定状态

19　在解释为什么条件（1a）和（1b）不能得到满足时，交易成本和信息成本显然非常重要。

20　Cass and Stiglitz［参考文献 3］。

21　假设每一组中的所有个人都是相同的。

22　为了某些目的，将两个企业看作是两个行业可能更好，因为我们要假设它们要相互竞争。

23　就目前而言，我们不用详细地说明最初的所有权（ownership claim）如何在第一家企业的股东之间分配。

24　假设在第二个行业中有投资。

25　如在［参考文献 24］中所设定的。

26　我们相应地省略了 X_1（θ）、$\hat{r}_1$ 等记号中的下标 1。

的发生概率，两组人的意见不一致。我们定义 $F^i(X)$ 为 i 组人所认为的 X 的分布函数，另外，很方便地定义 $\pi^i(X) = 1 - F^i(X)$。因为在两期模型中，当收入少于应偿付的债务时，即 $X(\theta) \leqslant (1+\hat{r}(B)\ B)$（其中 $\hat{r}(B)$ 是第一家企业发行数量为 B 的债券时所支付的名义利率）时企业就会破产，[27]所以 $\pi^i[(1+\hat{r})B]$ 是 i 所认为的企业不破产的概率。

我们定义：

$$\mu^i_{X \geqslant Y} \equiv \int_Y^{\infty} X dF^i(X) \tag{1}$$

$\mu^i_{X \geqslant Y}/\pi^i(Y)$ 是 X 的均值（条件于 i 组认为 X 大于 Y）。换句话说，$\mu^i_{X \geqslant (1+\hat{r})B}/\pi^i(\ (+\hat{r})B)$是企业不破产时回报的均值。

为简单起见，我们假设两组人中的所有人都是风险中立的，并且所有人不能将股票卖空（sell shares short）。我们还假设，对于所有的 Y，$\mu^a_{X \geqslant Y} > \mu^b_{X \geqslant Y}$，即相比另一组人，第一组人（第 i 个企业的原始股东）对企业的回报有更乐观的预期。偿付债务后，股东的回报为：

$$\begin{matrix} X^i(\theta) - (1+\hat{r})B & 当 & X^i(\theta) > (1+\hat{r})B \\ 0 & 当 & X^i(\theta) \leqslant (1+\hat{r})B \end{matrix} \tag{2}$$

其中

$$\hat{r} \geqslant r^*$$

从 i 看来，回报的均值是

$$\mu^i_{X \geqslant (1+\hat{r})B} - (1+\hat{r})B\pi^i((1+\hat{r})B) \tag{3}$$

a 组人完全所有的权益 E 的价值是

$$E = \frac{\mu^a_{X \geqslant (1+\hat{r})B} - (1+\hat{r})B\pi^a((1+\hat{r})B)}{(1+r^*)} \tag{4}$$

现在，我们需要决定函数 $\hat{r}(B)$，即名义利率如何取决于企业的借债？第二组人是风险中立的。因此，所需要的只是（在第二组人看来）他们所得到平均收益与无风险资产的收益（r^*）一样。风险债券的回报是：

$$(1+\tilde{r}) = \begin{cases} 1+\hat{r} & 当 \quad X \geqslant (1+r)B \\ \dfrac{X}{B} & 当 \quad X < (1+r)B \end{cases} \tag{5}$$

27　参见上文关于破产的定义以及［参考文献 24］。

因此，令$\mathscr{E}X^b$表示 b 组人对 X 的期望。

$$\begin{aligned}\mathscr{E}\{(1+r^b)B\} &= (1+\hat{r})B\pi^b((1+\hat{r})B) + \mathscr{E}\{X^b \mid (X(\theta) < (1+\hat{r})B)\} \\ &= (1+\hat{r})B\pi^b((1+\hat{r})B) + \mathscr{E}X^b - \mu^b_{X\geqslant(1+\hat{r})B} \\ &= (1+r^*)B\end{aligned} \tag{6}$$

将式（4）和式（6）相加，我们得到

$$V = E + B = \frac{\mu^a_{X(1+\hat{r})B} \geqslant \mu^b_{X\geqslant(1+\hat{r})B} - (\pi^a - \pi^b)(1+\hat{r})B + \mathscr{E}X^b}{1+r^*} \tag{7}$$

如果两组人的期望相同，则对于所有的 B：

$$\pi^a((1+\hat{r})B) = \pi^b((1+\hat{r})B) \tag{8a}$$

以及

$$\mu^a_{X\geqslant(1+\hat{r})B} = \mu^b_{X\geqslant(1+\hat{r})B} \tag{8b}$$

这样，企业的价值减至：

$$V = \frac{\mathscr{E}X^b}{1+r^*} = \frac{\mathscr{E}X^a}{1+r^*} \tag{9}$$

因此，企业价值独立于债务—权益比率，这正是我们已经说明的（因为这是均值—方差模型的一种特殊形式）。

但请注意，当两组人的期望不一致时会怎么样。只要没有破产，

$$\mu^a_{X\geqslant(1+\hat{r})B} = \mathscr{E}X^a \tag{10a}$$

以及

$$\mu^b_{X>(1+\hat{r})b} = \mathscr{E}X^b \tag{10b}$$

因此

$$V = \frac{\mathscr{E}X^a}{1+r^*} \tag{11}$$

债务—权益比率再一次没有带来任何改变。但是，如果

对于某些 θ， $B > X^b/1 + r^*$

第二组人会认为企业有可能（以正的概率）破产。企业的价值相对于债务进一步增加的导数是：

$$\begin{aligned}(1+r^*)\frac{dV}{dB} = \{&\mu^{a'}_{X>(1+\hat{r}_1)B} - \mu^{b'}_{X>(1+\hat{r}_1)B} \\ &- (\pi^{a'} - \pi^{b'})(1+\hat{r})B - (\pi^a - \pi^b)\}\frac{d(1+\hat{r})B}{dB}\end{aligned} \tag{12}$$

在我们的假设下，第一组人比第二组人更乐观；因此，即使第二组人认为破产的概率为正，第一组人也可能认为没有破产的可能。特别地，我们令

$$X_{\min}^{i} = \min_{\theta} X^{i}(\theta)$$

如果

$$X_{\min}^{a} > (1+\hat{r})B > X_{\min}^{b}$$

那么当第二组人认为破产的概率为正时，第一组人认为没有破产的可能。因而，在 $B(1+r^{*}) = X_{\min}^{b}$ 点

$$\frac{dV}{dB} = (X_{\min}^{b} - (1+\hat{r})B)f^{b}(X_{\min}^{b})\frac{d(1+\hat{r})B}{dB} = 0$$

其中 $f(X_{\min}^{b})$ 是出现 $X = X_{\min}^{b}$ 的概率（第二组人的估计中）。但是

$$\frac{d^2V}{dB^2} = -f^{b}(X_{\min}^{b})\left(\frac{d(1+\hat{r})B}{dB}\right)^2 < 0$$

因此，若投资固定，当债务超过了债权人认为有破产可能的债务总值时，企业的价值就会随债务上升而下降。

在上文中，我们假设破产不涉及真实成本（real cost），但我们仍推导出，一方面，如果企业的债务—权益比率很高（以至于债权人认为破产的概率为正）时，企业的价值会下降。如果在模型中考虑这些真实成本，只会进一步加强我们的结论。另一方面，如果个人是风险厌恶的（而不是我们这里所假设的风险中立），那么“新证券”（风险债券）的产生可能会增加企业的价值。这并不是因为利率是债务—权益比率的增函数（企业发行的债券越多，其价值越低），而是因为债权人和债务人对破产概率的估计不一致。[28]对债务人可以借到的数目起到限制作用的是债权人利用他们自己的概率估算得出的、能使得他们无差别的、必须补偿的金额，这高于债务人利用他自己的概率估算出的从借款中得到的收益。

四、财务决策和真实决策的互动

在前一部分，我们假设风险企业的投资水平（产出）已被固定，并且仅仅将注意力集中在财务决策上。既然我们已经知道了财务政策可能会影响

28 在日本箱根（Hakone）的讨论中，讨论会的几位参与者认为，在我的分析中，破产的假设并不是关键的，关键的是名义回报率是债务—权益比率的增函数。我希望上述论证已经让读者相信，仅有后者是不够的（不是充分条件）。另外，我也没有发现令人信服的论证可以证明，除了上述考虑的因素之外，为什么名义利率会是递增的。

到企业的价值，很显然，财务政策可能会影响企业的投资水平。我们将证明，最大化每一美元投资的收入可以同时得到最优的债务—权益比率和企业投资的均衡水平，并且两个决策一般来说不能分离。另外需要注意的是，经济在生产上可能是没有效率的。

规模报酬不变。我们假设对第一个企业的规模报酬为随机不变的：投入翻倍可以使每一自然状态下的产出翻倍。我们进一步假设，每组人都有一个固定数额的资金 I^i，需要进行投资；我们关注这些投资的竞争性配置。第一个行业中的代表性企业需要决定：有多大比例的投资需要以债务的形式进行融资。从上一小节的分析可以明显看出，名义利率仅为这一比例的函数，因为，当债券和投资水平同时翻倍时，任一状态下债券的回报都不变。竞争性[29]的企业相信，通过规模翻倍，它可以使价值翻倍；所需资金的一部分由债券融资（其比例与原先的一样），剩余部分进行股权融资。因此，因为我们假设了规模报酬不变，竞争均衡要求企业的总价值等于它的投资水平。否则，企业会扩张或缩小规模。

然而，从我们现在的观点来看，债务—权益比率的选择是我们的主要关注点。企业希望最大化其权益所有者每一美元投资的期望回报 $\mathscr{E}e^i(\theta)$，其中

$$e^i(\theta) \equiv \frac{X^i(\theta) - (1+\hat{r})B}{E} = \begin{cases} \dfrac{x^i - (1+\hat{r})\alpha}{1-\alpha} & \text{如果} \quad x^i > (1+\hat{r})\alpha \\ 0 & \text{其他} \end{cases} \tag{13}$$

此处

$$\alpha = B/I \tag{14}$$

是由债务融资的投资比例；并且

$$x^i = X^i/I^i \tag{15}$$

是所投资的每美元的回报。不破产的概率是 $\hat{r}$ 和 α 的函数。我们像在第三部分（式（1））一样定义 $\mu^i_{x \geqslant (1+\hat{r})\alpha}$ 和 $\pi^i((1+\hat{r})\alpha)$：如果 $G^i(x)$ 是 i 估计的 x 的分布，那么债务—权益比率为 $\alpha/1-\alpha$、支付的名义利率为 $\hat{r}$ 的企业在 i 的估计中不破产的概率是：$\pi^i((1+\hat{r})\alpha) = 1 - G^i((1+\hat{r})$

29 存在不确定性时，竞争均衡的含义并不是模棱两可的，这在［参考文献 22］中有详细讨论。正如本文中指出的，一般说来，所采用的特殊方法不是完全令人满意的，因为它不能应用于不存在规模报酬不变（或者至少存在乘积不确定性）的情形，也不能应用于项目存在可选择性（存在其他可行的分布）的情形。

α)，并且

$$\mu^i_{x\geqslant(1+\hat{r})\alpha} \equiv \int_{(1+\hat{r})\alpha}^{\infty} x dG^i(x)$$

$\mu^i_{x\geqslant(1+\hat{r})\alpha}/\pi^i((1+\hat{r})\alpha)$ 是企业不破产时的平均回报。企业希望最大化：

$$\mathscr{E}e^a(\theta) = \frac{\mu^a_{x\geqslant(1+\hat{r})\alpha} - (1+\hat{r})\alpha}{1-\alpha}$$

即

$$\mathscr{E}e^a(\theta) = (1+\hat{r}) + \frac{d\hat{r}}{d\alpha}\alpha \tag{16}$$

满足式（16）的 α 的值 α^* 是 α 的最优值，并且

$$\frac{\alpha^*}{1-\alpha^*}$$

是最优的债务—权益比率。

我们对式（16）有一个很自然的解释：权益投资的平均回报（等于当权益和债务同比例增加时的边际回报）一定等于借债的边际成本（即增加债务融资比例时的边际成本）。

如果对于 b 组人来说，除了投资于第一家企业的风险债券之外，还可以购买无风险债券（投资于无风险行业中），并且 b 组人是风险中立的，那么 $\hat{r}$ 一定会使得持有第一家企业风险债券的回报均值等于（$1+r^*$）。风险债券的回报均值为 $\bar{r}^b$（在 b 组人的估计中），且

$$(1+\bar{r}^b) = (1+\hat{r})\pi^b + \frac{\mathscr{E}x^b - \mu^b_{x\geqslant\alpha(1+\hat{r})}}{\alpha} \tag{17}$$

$\hat{r}$ 必须使得

$$(1+\bar{r}^b) = (1+r^*) \tag{18}$$

因此

$$\frac{d\hat{r}}{d\alpha} = \frac{\mathscr{E}x^b - \mu^b_{x\geqslant\alpha(1+\hat{r})}}{\pi^b\alpha^2} \tag{19}$$

将式（15）和式（17）至式（19）代入式（16），并且令 $\gamma=(1+\hat{r})\alpha$，我们得到

$$\mathscr{E}x^a = (1+\hat{r}) + \frac{d\hat{r}}{d\alpha}\alpha = \frac{(1+r^*)}{\pi^b(\gamma)} \tag{20}$$

式（20）结合式（17）可以改写成

$$\alpha(1+r^{*})=\gamma\pi^{b}+\mathscr{E}x^{b}-\mu_{x\geqslant\gamma}^{b}$$

这给出了含有两个未知数（γ，α）的两个方程。因而，我们可以解出均衡的债务—投资比率 α（或者是债务—权益比率 $\alpha/1-\alpha$）以及债券的名义利率 $1+\hat{r}=\gamma/\alpha$。

一个数值例子可能会很有用。假设 b 相信 x 在［1，2］上服从平均分布。那么

$$\pi^{b}=2-\gamma,\gamma\geqslant 1$$

$$\mathscr{E}x^{b}-\mu_{x\geqslant\gamma}^{b}=\frac{\gamma^{2}-1}{2},\gamma\geqslant 1$$

令 $\mathscr{E}x^{a}=1.67$，$1+r^{*}=1.5$。那么

$$\alpha=0.73$$

$$\gamma=1.1$$

$$(1+\hat{r})=1.51$$

即均衡中78%的投资由债务融资，破产的概率是10%，名义利率为51%。

到目前为止，我们一直假设第二组人总是把资金投资于无风险资产，或者用于购买第一家企业的债券。但是，他们同样也可以购买第一家企业的股份。如果第一家企业的股份的回报均值少于风险资产的回报或第一家企业债券的回报，那么他们就会购买股份。第二组人所认为的第一家企业的股份的回报均值是：

$$\mathscr{E}e^{b}=\frac{\mu_{x\geqslant(1+\hat{r})\alpha}^{b}-(1+\hat{r})\alpha\pi^{b}}{1-\alpha}$$

减去式（17），我们可以得到（做了一定简化后）[30]

$$\text{当}\ 1+\bar{r}^{b}\geqslant\mathscr{E}x^{b}\ \text{时},1+\bar{r}^{b}\geqslant\mathscr{E}e^{b}$$

这样，b 组人是否更偏好于投资于企业的风险债券而不是股份，这仅取决于在 b 组人的判断中，债券的回报是否高于企业每一美元投资的总回报均值。

现在，我们已经可以描述经济中所有的均衡了。第一，如果 $\mathscr{E}x^{a}>1+r^{*}$

30　如果企业发行的债券数目很大，以至于它的股东都相信有破产的可能，那么式（16）变为：

$$\mathscr{E}e(\theta)=-(\mu_{x>(1+\hat{r})\alpha}^{a'}-1)\frac{d(1+\hat{r})\alpha}{d\alpha}$$

其解释与上文中一样。

（无风险资产的回报），[31]第一组人的所有资本显然都投资于第一个行业的企业权益；如果 E^a 是 a 在第一个行业中企业的权益，I^a 是 a 要投资的资金，那么

$$E^a = I^a$$

第二组人会怎么做取决于 $\mathscr{E}x^b$ 和 $1+r^*$ 之间的关系。

（1）如果 $\mathscr{E}x^b > 1+r^*$，那么第二组人的所有资源也会都投资在第一个行业中。这可以是债券，也可以是股份。这存在两种情形。定义 α^* 为式（16）的解，其中 r（α）由式（17）给出，而且 $1+\bar{r}^b = \mathscr{E}x^b$（即第二组人中的投资者一定会在持有债券和权益之间无差异）：

（a）如果 $\alpha^* I^a/1-\alpha^* < I^b$，那么 b 组人所持有的风险债券 B^b 会满足 $B^b = \alpha^*(I^a + I^b)$。那么，对权益的投资额会是（$1-\alpha^*$）（$I^a+I^b$）。

（b）如果 $\alpha^* I^a/1-\alpha^* > I^b$，那么回报均值为 $\mathscr{E}x^b$，对借款会有过度需求，这造成债券的回报均值上升。这样，均衡就要求第二组人将所有资本投资于第一行业的债券，所获得的回报均值（$1+\bar{r}^b$）恰好介于两组人所认为的权益期望回报之间，即：[32]

$$\mathscr{E}x^b < 1 + \bar{r}^b < \mathscr{E}x^a$$

（2）另一方面，如果 $\mathscr{E}x^b <$（$1+r^*$），那么第一个行业中的投资由下式给出：

$$I^a(1+\alpha^*)$$

其中，α^* 又是从式（16）中得出，$\hat{r}(\alpha)$ 由式（17）和 $\bar{r}^b = r^*$（无风险的回报率）得出。与前面一样，这一结果是条件于 $\alpha^* I^a/1-\alpha^* < I^b$ 的。$\alpha^* I^a/1-\alpha^* > I^b$ 的情形也可以通过上述方法进行讨论。[33]

这里需要观察到的一个要点是：资源的真实配置和企业的财务政策是相互依赖的。在描述后者之前，我们不能解出前者。在规模报酬递减的情况下，这种依赖性更为明显。

规模报酬递减。现在，我们假设风险企业的回报由下述函数描述：$X(\theta) = h(I)\ \varphi(\theta)$，其中 $h' > 0$，$h'' < 0$ 并且 $\mathscr{E}\varphi = 1$，即存在乘积不确定性，但是规模报酬递减。当 $X(\theta) \leqslant (1+\hat{r})B$ 时，破产出现，即当

31　如果这一假设未得到满足，那么 a 和 b 都投资于无风险行业，这一问题也不再令人感兴趣。

32　确切的值可以通过下述方法获得。根据式（17），我们可以将 $\hat{r}$ 表示成 $\bar{r}$ 和 α 的函数，再将其代入式（16），并令 $\alpha^* = I^b/I^a$，我们就可以找到满足式（16）的 $\bar{r}$ 值。

33　即如果（$\alpha^* I^a/1-\alpha^*$）$> I^b$，那么 $1+r^* < 1+\bar{r}^b < \mathscr{E}x^a$。

$\varphi(\theta)\dfrac{h(I)}{I} \leqslant (1+\hat{r})\ \alpha$时。

如果$h(I)\ /I$是常数（存在规模报酬随机不变），破产的概率仅依赖于$(1+\hat{r})\ \alpha$，但是在规模报酬递减的情形下，给定任何债务—权益比率和名义利率，破产的概率随企业的规模递减。其结果是债券的名义回报率会反过来同时依赖于债务—权益比率和企业的规模，企业在做投资决策时，也会将这些考虑在内。最终的结果是，一般来说，这一经济在生产上不是有效率的。换句话说，如果第一个行业中存在几个有不同生产函数的企业，那么，

$$X^j(\theta) = \varphi(\theta)h^j(I) \tag{21}$$

其中，j表示第一个行业中的第j家企业。生产有效率要求：

$$h'^i = h'^j \tag{22}$$

对于该行业中的所有i，j都成立。但市场均衡一般不会满足式（22）。

为得到这一点，我们观察到，在均衡中，该行业中所有企业每一美元权益投资的回报必须一样都是：

$$\frac{\mathscr{E}(X^j)^a - (1+\hat{r}^j)B^j}{E^j} = \rho$$

这对所有的j也必须成立。假设企业的原始股东希望企业最大化他们在第一期的净财富：

$$\max_{[B^j,I^j]} \frac{\mathscr{E}(X^j)^a - (1+\hat{r}^j)B^j}{\rho} - (I^j - B^j)$$

其中I^j是投资的价值，因此$I^j - B^j$是通过发行新的股份而筹集的新资本。通过选择I^j和B^j，使得：[34]

$$\frac{\partial(\mathscr{E}X^j)^a}{\partial I^j} = \rho + B^j\frac{d\hat{r}^j}{dI^j} \tag{23}$$

$$\rho = (1+\hat{r}^j)\left[1 + \frac{B^j}{1+\hat{r}^j}\frac{d(1+\hat{r}^j)}{dB^j}\right] \tag{24}$$

式（23）和式（24）必须联立求解：真实投资决策不能与财务决策分离。

明确解出企业必须支付的利率$\hat{r}$，将其表示成投资水平和债务水平的函数（正如我们在上一小节所作的那样），然后将这些结果代入式（23）和式

34　将式（23）和式（24）结合到一起，意味着企业同样也最大化其股东每一美元投资的回报。

(24)，我们可以证明，式（22）一般来说不能得到满足。其中的细节留给读者完成。

五、收购

关于经理和股东。在这之前的讨论中，我们假设了投资水平决策是唯一的非财务决策。现在，我们将考虑，如果存在项目的多种选择（不同的生产技术、生产不同的产品等），结果会如何？如果对于选择何种项目存在意见不一致，又会怎样？我们关注的是由此导致的被收购的可能性，以及这对企业财务政策的意义。为了让企业做出一个决策，它必须确定：（1）不同状态的概率；（2）不同行动对企业市场价值的影响；（3）可供选择的可能行动。[35]

关于这些问题，职业人员（特别是经理）有可能比一般的股东拥有更多的信息。其结果是，股东很难判断经理是否做出了“正确”的决策。如果项目失败，这是不是因为运气太差，而经理对概率的估计是正确的？又或者是经理对项目结果的分布估计失误？又或是经理已经在可选范围内选择了最好的项目，但是所有项目的回报都很低？又或是经理不再追求价值最大化，而是有其他的目标（例如，他对企业规模更感兴趣）？

如果我们可以对相同的实验做重复尝试（这个世界明年的状态会和今年的一样），企业面临的机会都一样，经理做出了同样的决策，但是，像前面一样，结果取决于掷骰子（决定自然状态），这样，我们可以区分“坏运气”和“不称职”：平均来看，运气不佳的经理所在企业获得的回报会高于不称职经理所在的企业。不幸的是，我们不能做重复实验。不仅企业可选的机会集会发生变化，而且（即使机会集不变）上一次尝试的结果也会影响到我们对可选的机会集结果的分布的估计。即使经理的历史记录很好，当出现不好的结果时，股东也需要判断这仅仅是因为运气太坏，还是该行业的投资机会已经“逐渐枯竭”，又或者是经理不再是“有竞争力的”（因为环境变化，或者是经理个人发生了“变化”——如因为年纪变大）？

这一分析可以得到两个重要的结论。第一，与传统的新古典分析（新古典分析中的企业仅拥有资本和土地，以及一本“蓝皮书”，企业唯一的工作是在给定的要素价格下找到“蓝皮书”中正确的页码）不同，在这里，

35　如果存在阿罗—德布鲁证券的一个完全集，只要我们知道阿罗—德布鲁价格，我们可以很容易地确定对市场价值的影响，并且不用注意各种状态的概率。

企业家有至关重要的作用。例如，“蓝皮书”的内容可能只有他知道。尽管今天的要素价格是给定的，他必须猜测下一期要素的价格。因此，企业的市场价值不仅仅是其要素的市场价值（除了经理之外），其价值还取决于市场对经理会如何处理这些要素的估计。

第二，个人不仅对企业可选的不同项目的结果概率分布判断不同，而且他们对于经理做出正确（“合意的”）决策的能力的评价也不同。事实上，股东对于企业决策的不满意不会导致股东直接指示企业采取某种行动，相反，他们总是在那更换管理层。在实践中，通常不是股东直接鼓动更换管理层，而是另一被提名的管理层在鼓动更换。[36]

*收购的决定因素。*如果你拥有一家企业，而我观察到你的决策是错误的（在我看来），即平均来看，你从给定资本水平获得的回报低于我认为你可以获得的回报水平（暂时忽略风险厌恶），这样，我可以出价购买你的企业。如果在经济中普遍存在这种收购行为，那么所有的企业都会由（在他的判断中）能使企业获得最高回报的个人管理。然而，这一过程中至少有三个方面的局限性：

（1）就算我相信我在平均水平上可以比现在的管理层获得更高的回报，我的估计也会比管理者的估计有更大的方差，因为管理者有更多的“信息”，因而，我不会采取收购行为。用不正式的话讲，尽管我相信我可以在平均水平做得比现在的管理层好，但我“对此并不确信”。

（2）交易成本。这是对收购的相关成本的一个总称。其中一项重要的成本是：如果我将经理解雇，他可能会将一些重要的、拥有内部信息的公司员工带走；如果我要找人替代他们，这将花费很大的成本。

（3）我可能没有足够的资本，来购买足以控制这个企业的股份。

企业现有的股东关注收购，至少有两个原因：

（1）收购的可能性有可能会使得企业应支付的债券利率升高。当初始的债券持有者借钱给企业时，不仅会考虑到在现有管理层下企业在各种状态下可以获得的回报，而且也会考虑到潜在的收购——新的管理层会采用一系列新的政策。他们必须考虑所有可能的收购，这些收购所能导致的企业的回报差距可能非常大。这样，原始债券持有者所要求的回报不仅取决于当前管

---

36 对此有几个很好的理由：（1）判断特定的管理层的能力好坏，这是一个相对的问题；每个股东都认为某个管理层可以做得更好，因此它是“差”的，但在不同的管理层中，它可能是最好的；（2）需要管理层自身来组织一项收购行动。

理层管理下的破产概率，也取决于在其他管理层管理下的破产概率，以及其他管理层出现的概率。

（2）尽管企业的大部分股东因收购而境况改善——否则他们不会出售其企业——少数股东的股价可能高于市价，他们享受着（过去也享受着）一些消费者剩余；而当新的管理层改变企业政策时，他们将失去这些剩余。因此，他们可能因收购而境况变差。

企业可能希望能避免收购，这一点似乎是很清楚的。问题是，他们可以采取怎样的行动使其被收购的可能性减少。回到我们所列举的收购的局限性上，我们发现可以通过以下几种方式减少被收购的可能：

（1）采取一些不大可能让其他管理层认为还有可以做得更好的行动。具有事前“风险”的行动总会有失败的风险，因此，企业总是倾向于采取一些更保守的行动——如果不存在收购，企业会采取更具风险性的行动。[37]

（2）采取一些行动，使得希望收购它们的个人或团体需要借更多的钱来收购，从而降低了他们借款条件的吸引力，也降低收购后能从企业获得的回报。此处有三点相关的考虑：

（a）给定市场估价，债务—权益比率越大，企业越有可能被收购。要取得该企业的控制权，个人仅需积累足够的资本，以购买企业大部分的股份。因此，对于某个给定市场价值的企业来讲，债务—权益比率越大，就越容易被收购。

这一论断有一重要的问题：如果个人贷款[38]是企业贷款的很好的、合理的替代品，那么企业的债务—权益比率越小，个人可以借到的资金越多——这正好抵消了企业借款的缺乏。然而，有两个原因使得事实与此相反。非常明显的一个原因是：管制限制了个人用证券作为抵押品时的贷款数额。第二，债券持有人可能要求个人支付比企业债券高的利率；并且，他们要求的利率可能是贷款数额的增函数。此外，个人可能还有别的债务，因此在某些状态下，个人会破产，而企业不会。[39]当然，也存在另一事实，可以抵消这一效果——个人可能拥有其他资产，因此，即使企业破产了，他也可能不会破产。另外，债权人对企业破产概率的估计取决于谁在管理这家企业，因为

37　这一情形不能太极端化，因为那样的话，其他管理层可能认为，一些“良性风险”被忽略了。

38　因为许多情况下是由其他企业对某一企业进行收购，因此下列论点更站得住脚：我们用进行收购的企业的直接借贷替代了被收购企业的“间接”借贷。

39　因此，个人可能确实不愿意借很多资金，除非存在上文中所讨论的有限责任的某种安排。

只有管理权的控制权才能产生真正的影响。管理层对于试图收购的个人会要求更高的名义回报率（与原始管理层相比），因为他们对其将对企业所做的决策了解得更少（即使他们对收购者业绩良好有很高的“期望”，但他们对结果的预期分散度更高，因此他们对破产的预期也更高）。当然，试图收购的个人很有可能有很好的声誉，因此他在任何贷款水平上都有可能比企业贷款便宜，这有助于收购时的出价。[40]

（b）如果在收购者的眼中，贷款的名义利率比收购后（在收购者计划的债务水平上）可得的利率低，那么较高的债务—权益比率可以增加收购的吸引力。另一方面，如果贷款的名义利率比收购后可得的利率高（收购者的估计），那么较高的债务—权益比率将减少收购的吸引力。[41]

（c）很有可能企业越大，就越难被收购。考虑有两个企业，其中一个企业的规模是另外一家企业的两倍，而其他所有方面都一样（包括债务—权益比率）。试图收购较大企业的集团必须调动更多的资源，特别是要借到更多的资金。给定收购者必要债务的一个百分点的增加，它的总资产也将增加一个百分点（假设像通常的那样，进行收购的企业或个人最初拥有一定的资产）；[42]需要借的资金越多，在借款人看来，风险也越大，他们所要求的利息也就越高。但是，在上述两种情形下，可能正好与一般的假设相反。如果在借款人看来，被收购的企业的回报风险性低于进行收购的企业（即使考虑了新管理层所带来的不确定性），那么，当被收购的企业越大时，进行收购的企业可以借到的利率也越低，这个投资机会也就更加吸引人。这正是一个小企业试图收购大企业时的典型情形。

从我们的分析中可以看出，没有一种政策可以避免所有被收购的可能，某种行动在阻止了某一组人（对他们的借贷有特定的期望、拥有一系列的资源、拥有一系列的生产计划）出价收购的同时，鼓励了另一组人的出价收购。然而，这里好像有一个假设：一般来说，可以增加企业权益价值的行动可以使得收购出价更加不容易掌控。但是，即使在这里，这个假设也是模棱两可的。企业当前管理层认为可以使下一期市场价值很高（概率也很高）

---

40 事实上，美国大集团的很多成功收购都需要大规模的借贷，好像市场对这些经理有很大的信心。随后的事件令人怀疑这些信心是否在事实上是正当的。

41 必须区分下列两种利率：收购集团为了能成功收购而必须支付的利率（涉及进行收购的分司的债务），以及被收购的企业在收购后可以借贷的利率。

42 此处相关的资产估价是贷款到期时债权人对资产价值的评估。进行收购的企业的净资产（从这一点来看）可能是负的。

的行动，使得下一期不大可能被收购，但也有可能使得这一期的市场估价降低，从而使得这一期更有可能被收购。

一个例子。下述例子对上文的例子做了很小的改动，证明了被收购的威胁如何影响企业的行动。

假设存在三组人。每一组内的人都是完全一致的，并且是风险中立的。a 组人最初拥有行业 1。行业 1 有两个可选的项目，两个项目都是规模报酬随机不变的，但是它们所带来的结果服从不同的概率分布。在 a 组人的估计中，项目 a 的回报比无风险资产的回报高很多；但在 b 组人的估计中，它的回报期望为零；在 c 的估计中，它的回报为正，且稍高于 r^*；在 a 组人和 c 组人的估计中，项目 b 回报为零；但在 b 组人的估计中，它的回报非常高。

首先，我们考虑 a 组人和 c 组人忽略收购的可能性时的市场均衡。这在实质上与第三部分的讨论一致。项目 a 会被采用。b 组人会投资于无风险产业。

第一家企业的债务—权益比率将是：

$$\frac{\alpha^*}{1-\alpha^*}=\frac{\gamma I^c}{I^a+(1-\gamma)I^c} \tag{25}$$

其条件是（正如我们应假设的）

$$\frac{I^c}{I^a}\geqslant\frac{\alpha^*}{1-\alpha^*} \tag{26}$$

其中 γ 是 c 以债券形式拥有的投资比例。但是，如果考虑收购的可能性，在下述条件成立时，b 组人会收购该企业：

（1）他们的期望回报大于 r^*；因为他们不可能发行债券，这意味着：

$$\frac{\mu^{bb}_{X_1>(1+\hat{r}_1)B_1}-(1+\hat{r}_1)B_1\pi^{bb}_1((1+\hat{r}_1)B_1)}{E_1/2}>1+r \tag{27}$$

其中 $\mu^{bb}_{X_1>(1+r_1)B_1}$ 是 b 组人对行业 1 的条件期望利润（如果项目 b 被采用）。请注意，为了获得企业的控制权，只需要收购一半的股份；但是，一旦企业获得了控制权后，股东手中的另一半股份在 a 组人看来是毫无价值的，因此，如果 b 组人合谋行动，这些股份可以毫无成本地获得。

（2）如果他们收购这一企业是可行的，即他们的总资源大于企业价值的一半：

$$\frac{E_1}{2}<I^b \tag{28}$$

从式（25）可知，权益的价值为

$$I^a + (1-\gamma)I^c$$

这有可能大大少于 I^b，保证了收购是可行的；并且，如果 $\mu^{bb}_{X_1>(1+\hat{r}_1)B}$ 的值足够大，那么收购是非常值得的。

尽管收购后，b 组人会因此而境况改善；但 a 组人和 c 组人可能会因此境况变差：

（1）c 组人的债券（在它看来）是毫无价值的；

（2）尽管 a 组人中有一半的股东对于新的安排无所谓（否则他们不会出售他们的股份），但另一半人将遭受很大的资本损失（因为在他们看来，股份不能再带来任何回报，b 组人也可能不会在股份原有价格下购买所有的股份）。因此，从整体来看，a 组人的境况有所恶化。

因此，正如我们已经论证的，债券持有人和股东都会将破产的可能性考虑在内。在我们的例子中，这将以下述的形式出现。如果 E_1 小于 $2I^b$，c 组人会拒绝出借任何东西。这意味着利率函数存在明显的不连续性。另外，不像前一例子那样（名义利率仅为债务—权益比率的函数），这里的利率是债务—权益比率和权益绝对水平的函数。[43]这样，此例中均衡的债务—权益比率不由式（25）给出，而是由下式给出：

$$\frac{I^a + I^c - 2I^b}{2I^b} \tag{29}$$

对收购的进一步评论。在上述讨论中，我们可以回避了两个领域的问题。第一个领域的问题与竞争性市场均衡是否存在相关，其中收购可能是很重要的。事实上，因为"控制权"依赖于51%的股东，所以，"竞争均衡"的确切含义值得商榷。但是，尽管很容易构建市场均衡的扩展定义，考虑到存在股市经济中的非市场因素（股东大会，或者是收购的可能性——购买50%的股份），没有什么可以确保存在这样的均衡。考虑市场均衡的最低要求：

（1）在所有可能的计划中，大多数原始股东（以股份加权）一定偏好于给定的计划。

（2）在该计划被宣布后，这个计划必须被大多数股东所偏好：既然他们已经购买了该股票，要重新举行股东大会来采用新的计划，这肯定是不合算的。

43　回忆在第四部分中，我们已经假设企业间相互竞争，即我们假设：如果企业将债务和投资翻倍，它的市场价值也会翻倍。当时，我们已经表达了对此的一些疑虑；既然权益的水平和债务—投资比率都会影响利率，我们对这个假设必须存有更多的疑虑。

（3）收购大部分的股份并采取一个新的计划，这对任何个人（不管是原始股东，还是计划宣布后的股东，甚至根本不是股东的人）都是不合算的：成立一家“控股公司”，用来购买给定企业的股份以获得对该企业的控制权，并且承诺在获得控制权后采用新的计划——这对任何企业家都是不合算的。

任何计划可能都不能满足这些要求，原因如下：

（1）除非在非常限制性的条件下，我们在第一个要求和第二个要求中会碰到一般的投票问题：可能不存在一个计划，可以相对于其他所有计划得到大多数人（循环的大多数）的偏好。即使经济中仅有两组人，可能也很困难。

（2）例如，我们考虑这样一种情形：企业有两种生产计划；其中之一在一组人的估计中有很高的回报，但在另一组人的估计中回报很低；另一个生产计划的情形正好相反。因为只需要“收买”大半的股东，很明显，如果企业原先由第一组人控制，那么第二组人可以“收买”第一组人中的大多数；如果企业由第二组人控制，那么第一组人可以“收买”第二组人的大多数。因此，第三个要求不能得到满足。仅当合谋行动时，这样的收购才能被阻止。

（3）事实上，即便不正式的收购也会遇到这些问题。在上述例子中，如果企业最初由第一组人拥有，由于项目的期望回报（同时方差很小）比较高，因此个人的期望效用很高。因此，与回报很低（方差相同）时个人会投资于无风险企业的份额相比，个人会将其资产的很大一部分投资于无风险企业。很明显，如果第一组人最初拥有该企业大多数股份，并采用他们偏好的项目，那么结果很有可能是最后他们只占一小部分股份；如果第二组人最初拥有大多数股份，情形正好相反。这样，第一个要求和第二个要求可能不能被同时满足。

（4）如果多于一种要素或商品，那么就会出现第四个问题。为了建立均衡，不同商品和要素的过度需求曲线必须是连续的（或是上半连续的）。本期的生产取决于下一期不同自然状态下的期望价格，因此也取决于下一期的生产计划。通常，我们认为本期价格的微小变动对下一期的生产计划和期望价格有很小的影响，因而对本期采用的生产计划也仅有微小影响——因此，对本期的过度需求也仅有微小影响。但是，如果价格的微小变化导致了企业从利润和应付债务相等变到了利润少于应付债务（企业破产了），那么新的管理层——他们对其他项目未来的期望回报与原来的管理层有很大差别——所采用的决策与先前价格下的决策可能有天壤之别。但是，这意味着

我们没有任何理由说，本期的过度需求曲线是连续的。

在这里，我们仅可隐含地表明，构建市场经济均衡存在性所涉及的问题正是现代资本主义经济的一些特征。[44]我曾考虑了一些非市场的均衡概念，特别是核（the core），但是它们同样是不能令人满意的。事实上，我们可以给出一些核为空的例子。

第二个领域的问题涉及信息、期望回报的决定，以及企业的价值。个人持有企业股份的回报不仅包括红利，也包括资本利得。因此，本期的价格取决于下一期的期望价格。下一期的期望价格并不怎么依赖于我们所认为的企业回报在下一期的贴现值，而是依赖于市场如何评估其价值。很明显，这有可能导致投机性繁荣——这在其他文献中已有讨论。那些文献已经证明：（在确定性模型中）“每次投资性繁荣一定会结束”。在有限的时间内，价格一定会变成无限；在泡沫破灭时，一些人会最终持有这些股份。可能有人会认为，市场可以“学会”如何避免这样的行为；但是，如果不存在期货市场，好像不存在什么机制可以避免这些问题。这里我想指出的是，当存在不确定性时，可能出现进一步的问题。若股票不能带来期望的高回报时，个人必须就下述情形做出判断：（1）他只是运气欠佳——他早已认识到，在某些自然状态下，价格会很低，并且对这些状态赋予了较低的概率，然而，这些状态有时还是发生了；（2）信息得到了进一步的披露，因此影响了其他人对未来回报的期望，从而也影响了股票的价格；或者（3）最初的价格“太高”了。不存在确切的理由可以区分这三种情形。

个人必须部分依赖于别人对企业价值的估计，这一事实对于收购有两个重要的意义。

第一，个人或者一组人的估计（控制企业所能得到的价格高于企业的现价）对股东是一种信息；结果有可能是，他们会调高对企业回报的估计。事实上，如果企业或者个人的收购出价仅稍高于市场的现价，股东们可能会因此调整他们估计的市场价值，并且高于收购的出价。这至少部分地解释了为什么收购的出价会大大高于当时企业的市场价格。

第二，在借款时，借款人可能对破产概率知之甚少。如果他观察到很多想法类似的人要求的利率高于无风险回报，他可能认为，这表明了他人认为破产的概率很大。因此，他可能改变他对破产概率的估计，要求以一个更高的名义回报。这些事实使得可贷资本的供给曲线在超过某一点时变得非常陡

44 另外，可能同等重要的是，不再利用阿罗—德布鲁证券（事实上，这种证券并不存在）。

峭，使得个人或企业要借款筹资收购其他企业时难度更大。

六、结论

本文已经证明，破产的可能性对企业行为有重要的意义：我们不能将财务决策和真实决策相分离，同时也确实存在一个最优的债务—权益比率。另外，我们证明了更进一步的后果——经济可能在生产上是无效率的。

一些经济学家认为，因为破产在经济中并不常见，所以它不可能是重要的；一个假设破产不可能出现的模型比本文给出的、破产有重要作用的模型更"真实"。我们要论证的是，就算破产是相对少见的，破产的可能性也会对企业的行为有重要的影响；例如，他们会选择合适的债务—权益比率，以避免较高的破产概率。另外，我们应该将收购和兼并与破产归于一类；濒临无力偿债的企业更偏好用收购和兼并来渡过难关，而不是承担很高的成本实行真正的破产。近年来，这些现象在美国和英国都非常普遍。

我们认为，在本文研究的模型中，不仅这里的假设比传统模型中的假设更真实，模型的预测也与真实企业行为更为一致。不管怎么讲，包含破产和不一致预期的模型需要我们做更多的研究。

参考文献

1. Arrow, K. J. "Aspects of a Theory of Risk-Bearing." Yrjo Jahnsson Lectures, Helsinki, 1965.
2. ———. "The Role of Securities in the Optimal Allocation of Risk-Bearing." *Review of Economic Studies,* Vol. 31, No. 2 (April 1964), pp. 91–96.
3. Cass, D. and Stiglitz, J. E. "The Structure of Preferences and Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds." *Journal of Economic Theory,* Vol. 2, No. 2 (June 1970), pp. 122–60.
4. Debreu, G. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium.* New York: John Wiley & Sons, Inc., 1959.
5. Diamond, P. A. "The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty." *The American Economic Review,* Vol. 57, No. 4 (September 1967), pp. 759–76.
6. Hahn, F. H. "Equilibrium Dynamics with Heterogeneous Capital Goods." *Quarterly Journal of Economics,* Vol. 80, No. 321 (November 1966), pp. 633–46.
7. ———. "Equilibrium with Transactions Costs." *Econometrica,* Vol. 39, No. 2 (May 1971), pp. 417–40.
8. Hirshleifer, J. "Investment Decision under Uncertainty: Applications of the

State-Preference Approach." *Quarterly Journal of Economics,* Vol. 80, No. 2 (May 1966), pp. 252–77.
9. ———. "The Private and Social Value of Information and the Reward to Inventive Activity." *The American Economic Review,* Vol. 61, No. 4 (September 1971), pp. 561–74.
10. JENSEN, M. AND LONG, J. B., JR. "Corporate Investment under Uncertainty and Pareto Optimality in the Capital Markets." *The Bell Journal of Economics and Management Science,* Vol. 3, No. 1 (Spring 1972), pp. 151–74.
11. KNIGHT, F. H. *Risk, Uncertainty and Profit.* Clifton, N. J.: Augustus M. Kelly, 1921.
12. LINTNER, J. "The Aggregation of Investor's Diverse Judgements and Preferences in Purely Competitive Security Markets." *Journal of Financial and Quantitative Analysis,* Vol. 4, No. 4 (December 1969), pp. 347–400.
13. ———. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *The Review of Economics and Statistics,* Vol. 47, No. 1 (February 1965), pp. 13–37.
14. MODIGLIANI, F. AND MILLER, M. H. "The Cost of Capital, Corporation Finance and the Theory of Investment." *The American Economic Review,* Vol. 48, No. 3 (June 1958), pp. 261–97.
15. MOSSIN, J. "Equilibrium in a Capital Asset Market." *Econometrica,* Vol. 34, No. 4 (October 1966), pp. 768–83.
16. RADNER, R. "Competitive Equilibrium under Uncertainty." *Econometrica,* Vol. 36, No. 1 (January 1968), pp. 31–58.
17. ROTHSCHILD, M. AND STIGLITZ, J. E. "Increasing Risk: A Definition and Its Economic Consequences." *Journal of Economic Theory,* Vol. 2, No. 3 (September 1970), pp. 225–43, and Vol. 3, No. 1 (March 1971), pp. 66–84.
18. SHARPE, W. F. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance,* Vol. 19, No. 4 (September 1964), pp. 425–42.
19. SHELL, K. AND STIGLITZ, J. E. "The Allocation of Investment in a Dynamic Economy." *Quarterly Journal of Economics,* Vol. 81, No. 325 (November 1967), pp. 592–609.
20. STIGLITZ, J. E. "The Degree of Product Differentiation in Monopolistic Competition: An Example." Mimeographed. The Cowles Foundation, Yale University, 1972.
21. ———. "On the Irrelevance of Corporate Financial Policy." Cowles Foundation Discussion Paper No. 339, Yale University, April 1972.
22. ———. "On the Optimality of the Stock Market Allocation of Investment." *Quarterly Journal of Economics,* Vol. 86, No. 342 (February 1972). pp. 25–60.
23. ———. "Perfect and Imperfect Capital Markets." Paper presented at the meeting of the Econometric Society, New Orleans, 1971.
24. ———."A Re-Examination of the Modigliani-Miller Theorem." *The American Economic Review,* Vol. 59, No. 5 (December 1969), pp. 78–93.
25. ———. "Taxation, Corporate Financial Policy, and the Cost of Capital."

Journal of Public Economics, forthcoming.
26. ______. "Value Maximization and Alternate Objectives of the Firm." Mimeographed. The Cowles Foundation, Yale University, 1972.
27. STIGUM, B. "Competitive Equilibrium under Uncertainty." *Quarterly Journal of Economics,* Vol. 83, No. 333 (November 1969), pp. 533–61.
28. TOBIN, J. "Liquidity Preference as Behavior toward Risk." *Review of Economic Studies,* Vol. 25, No. 1 (February 1958), pp. 65–86.

股市投资配置的最优性 *

一、引言

早期关于无风险投资（safe investment）和风险投资（risky investment）的市场配置问题的很多研究认为，市场会系统性地分配极少的资源给风险投资，因为许多私人的风险并不是社会风险。然而近期的文献却表明，本质上股票市场提供了所有可能被提供的风险分担的可能性（risk - pooling possibilities），“股票市场构建了一个相当有效的配置机制。”[1] 如果假设存在一个或有商品（contingent commodity）的完全集合（complete set）（即有多少种自然状态，在股票市场上就至少有多少种证券），上述论断完全成立；这一点已经由阿罗和德布鲁[2] 给出了更为精确的证明。然而，这个假设却受到理论上和实证上的强烈质疑。[3] 如果证券数量少于自然状态的数

---

* “On the Optimality of the Stock Market Allocation of Investment”, *Quarterly Journal of Economics*, 86（1）, February 1972, pp. 25 - 60. 本文的早期版本，曾在1970年在东京举办的计量经济学协会远东会议上报告过。本研究得到了古根汉基金会、福特基金会以及美国国家科学基金会的赞助。作者在此需要感谢 A. 克莱韦瑞克（A. Klevorick）、S. 小泉（S. Koizumi）、P. 神谷（P. Kamiya）以及 M. 简森（M. Jensen）对本文早期版本的有益评述。M. 简森和 J. 朗（J. Long）在“On the Inconsistency between ‘Optimal’ Corporate Investment and Pareto Optimality”（mimeo, University of Rochester, 1969）一书中报告了与本文类似的结果。

1 W. J. Baumol, *The Stock Market and Economic Efficiency*（New York: Fordham University Press, 1965）.

2 K. J. Arrow, “The Role of Securities in the Optimal Allocation of Risk Bearing,” *Review of Economic Studies*, Vol. 31（April 1964）, pp. 91 - 96。以及 G. Debreu, *The Theory of Value*（New York: John Wiley & Sons, 1959）.

3 更详细的讨论可以在下列文章中找到：J. E. Stiglitz, “Some Aspects of the Pure Theory of Corporate Finance, Bankruptcy, and Take - Overs,” 在箱根（Hakone, Japan）做的报告, June 1970; J. E. Meade, *The Theory of Indicative Planning*（Manchester University Press, 1970）; R. Radner, “Competitive Equilibrium under Uncertainty,” *Econometrica*, vol. 36（Jan. 1968）, pp. 31 - 58; K. J. Arrow, Lecture 3, *Some Aspects of the Theory of Risk - Bearing*（Helsinki: Yrjö Johnssonin Säätiö, 1965）; F. H. Hahn, “Equilibrium with Transactions Costs,” *Econometrica*, vol. 39（May 1971）, pp. 417 - 440.

量，[4] 就会产生三个问题：

（1）我们如何评价其他配置方式？如果没有阿罗—德布鲁证券的一个完整集合，在不同状态下，不同人的消费边际替代率就不会总是相同。因此，将存在阿罗—德布鲁市场完整集合的经济体（这在传统意义上是帕累托最优的）和缺乏这些市场的经济体进行比较就不是特别有意义。[5] 我们给出如下标准：如果一个中央计划者（受约束于在一个给定的市场结构下买入要素并售出商品）无法在不使任何人福利下降的情况下使某人福利提高，这样的配置就是（受约束的）帕累托最优配置。[6]

（2）每个企业的目标是什么？传统的理论认为企业的目标是使其股市价值最大化；由于个人购买商品的价格不受企业行为的影响，通过增加股东的财富，每个股东的福利都能得到最大化。无论个人是企业的大股东还是小股东，这一点都是成立的。但如果没有阿罗—德布鲁证券的完整集合，价值最大化的假设就不成立。

（3）个人要评估其他生产计划，就必须确知生产计划对股价（企业的价值）的影响。在阿罗—德布鲁模型中，对应于每一种自然状态都有一个产出（投入）价格；企业的净价值就是以这些价格计算的产出价值和投入价值的差额。[7] 然而，如果没有一个阿罗—德布鲁市场的完全集合，企业如何评估其他的生产计划呢？

4 即使在不存在不确定性的情况下，这里所讨论的问题也有可能出现，如下述情形：不同商品所有可能被生产出的"质量"连续统（continuum）中，仅有很少一部分被真正生产出。风险资产市场的分析提供了一种场景（context）；在这个场景中，相比在一般情形下，这些问题的分析（在某种程度上）可以更清晰、更精确。另外，对于我们将要讨论的这些困难来说，这一条件——证券的数量少于自然状态的数量——既不充分也不必要，我们将在下文更清晰地讨论这一点。

5 这一点已经由鲍启（Borch）证明，可以参见他的下列文章："The Economics of Uncertainty" in M. Shubik, ed., *Essays in Mathematical Economics*, In Honor of Oskar Morgenstern (Princeton University Press, 1967), pp. 197 - 210.

6 这一概念等同于P. 戴梦德（P. Diamond）定义的"受约束的帕累托最优"（Constrained Pareto Optimal）——任何个人所分得的收入必须是经济中不同企业产出的一个线性函数。请参见：P. Diamond, "The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty," *American Economic Review*, vol. 57 (Sept. 1967), pp. 759 - 776.

7 还需明确一点：如果经济中存在的相互独立的证券的数目等于自然状态的数目，那就相当于在经济中存在阿罗—德布鲁市场的一个完全集合。请参见：D. Cass and J. E. Stiglitz, "The Structure of Preferences and Returns and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds," *Journal of Economic Theory*, vol. 2 (June1970), Appendix A, pp. 122 - 160.

本文主要关注第三个问题。[8] 我们将给出一个具体的例子，在该例中可以用一种简单且可被接受的方法来评估其他生产计划。在这一例子中，企业是竞争性价格接受者；但是，当企业使股市价值最大化时，并不能得到一个最优配置（如上文的定义）。更宽泛的意义上，我们将要通过此例说明，为什么企业在股市价值最大化时，其市场配置不可能是最优的。

二、基本模型

我们在详细探究的具体例子之前，需要对经济体的基本结构给出一个大致框架。该经济体由 $n+1$ 个企业（行业）[9] 以及 m 个个人组成。我们将考虑一个简单的两期模型，其中单一的投入要素由 I（投资）表示，它在第一期购得，并在第二期转化为单一的产出。[10] 企业（行业）之间的不同在于某些自然状态下的回报不同，而非生产的产品不同。[11]

1. 个人行为

每个人在“生命”开始时有一定的禀赋 I^j，并且对不同的公司有一系列股份。β_i^j表示在第一期开始的时候，第 j 个人[12] 对第 i 个企业拥有的股份比例（$\sum_j \beta_i^j = 1$）。如果企业在第一期不做任何投资，那么个人将获得企业在第二期 β_i^j比例的产出。当然，个人可以出售它对于第 i 个企业的股份，用以购买其他企业的股份。一般来说，我们可以预见个人会重新规划他的“投资组合”。另外，个人对第 i 个企业的初始股份的价值取决于该

---

8 关于第二个问题（企业的目标）的讨论，请参见：J. E. Stiglitz, “Value Maximization and Alternative Objectives of the Firm,” mimeo, Cowles Foundation, 1972，以及 J. E. Stiglitz, “Some Aspects of the Pure Theory of Corporate Finance,” 前引文献。

9 本文的大部分分析中，n 是固定的。在第五部分，我们将简单讨论 n 是内生决定时的情形。这些在下述文章中有更详尽的讨论：J. E. Stiglitz, “The Degree of Product Differentiation in Monopolistic Competition: An Example”。

10 只要使用更复杂的符号，我们就可以轻易地将模型扩展到多个投入的情形。如果考虑多个产出，模型中会引入价格的不确定性。毫无疑问，相对价格的不确定性是非常重要的，它也会导致其他一些问题，但本文将不对此展开讨论。相关的研究有：J. E. Stiglitz, “Behavior Towards Risk with Many Commodities,” *Econometrica*, vol. 37 (Oct. 1969), pp. 660 – 666; J. E. Stiglitz, “A Consumption Oriented Theory of the Demand for Financial Assets and the Term Structure of Interest Rates,” *Review of Economic Studies*, vol. 37 (July1970), pp. 321 – 351；以及 P. Diamond（前引文献）绪论中的介绍。

11 由于假设了只有一个商品，不确定性的唯一来源就是技术。参见脚注 10。

12 此处以及本文其他地方，我们采用上标 j 表示个人，下标 i 表示企业（在后文中，在不引起混淆的情况下，我们将省略上标和下标）。

企业的投资水平。如果第 i 个企业投资了 I_i，并且其全部市场价值为 V_i，[13] 那么原始股东的资产价值就是 $V_i - I_i$，即企业的全部价值减去其用于新的投资的资本。[14]莫迪格利安尼—米勒定理保证了企业的价值独立于它为新投资融资的方式；即它不取决于企业的债务—权益比例。[15]因此第 j 个人的初始财富 W_0^j 就是：

$$W_0^j = I^j + \sum_i \beta_i^j (V_i - I_i) \tag{1}$$

个人考虑可以获得的各种证券，[16]并且将自己的财富分配于这些证券用以最大化他在下一期的期望效用；也就是说，假如 $e_i(\theta)$ 是投资于第 i 个风险企业的股票的单位美元的回报，r 是投资于某一无风险证券（债券）（即在各种自然状态下回报相同的债券）的回报，并且假如个人花费了 V_i^j美元用于购买第 i 个企业的股票，[17]他在状态 θ 下的收入 $Y^j(\theta)$ 就是：

$$Y^j(\theta) \equiv \sum_{i=1}^{n} V_i^j e_i(\theta) + r\left(W_0^j - \sum_{i=1}^{n} V_i^j\right) \tag{2}$$

个人选择 V_i^j 以让下式最小化

---

13 我们的计价单位是 I。在模型中的所有经济行为（交易和投资决定）都发生在第一期；因此我们只关心企业在这一时期的价值决定。将这一分析扩展到多期（尤其是无穷期）范围有一些本质上的困难，包括企业当期的价值取决于它在下一期的期望价值，而其下一期的期望价值又进而取决于再下一期的期望价值，如此类推；在确定的情况下，这些问题已在下述文章中有所讨论：K. Shell and J. E. Stiglitz, "The Allocation of Investment in Dynamic Economy," *Quarterly Journal of Economics*, vol. 81 (Nov. 1967), pp. 592 - 609.

14 考虑下述不同的企业可能是很有意义的——不同的农场在不可分的土地上运营。农场的全部价值 V_i 就是土地的价值加上农场投资的价值 I_i。

15 F. Modigliani and M. H. Miller, "The Cost of Capital, Corporation Finance, and the Theory of Investment," *American Economic Review*, vol. 48 (June 1958), pp. 261 - 297; J. E. Stiglitz, "A Re-Examination of the Modigliani-Miller Theorem," *American Economic Review*, vol. 59 (Dec. 1969), pp. 784 - 793. 自始至终，我们假设没有交易成本，并且对于不同的融资工具不课以不同的税。这些都是非常重要的假设。参见 J. E. Stiglitz, "Taxation, Risk-Taking, and the Allocation of Investment in a Competitive Economy," in M. Jensen, ed., *Studies in the Theory of Capital Markets*。只要每个人在一种自然状态下对另一种自然状态下的边际收入替代率不同，就会刺激人们创造额外资产。但也有很多理由说明了为什么这种情况不发生。参见脚注 3。

16 我们明确地假定除了 $n+1$ 个企业提供的资产之外，没有其他资产（并且没有其他收入来源）。当然上述分析可以扩展到其他情况，如果在其他情况下也如看起来那样的确如此的话——也就是说证券数量少于自然状态数。

17 我们允许卖空（short sales）和借入，即：

$$V_i^j \leqslant 0, W_0^j - \sum_{i=1}^{n} V_i^j \geqslant 0。$$

$$Eu(Y^j(\theta)) \tag{3}$$

在此，像往常一样，我们假设个人是风险厌恶的（$u'>0$，$u''<0$）。因此个人对第 i 个企业证券的需求就是关于 $e_1, \cdots, e_n$，r，W_0^j 的一个函数：

$$V_i^j = D_i^j(e_1, \cdots, e_n, r, W_0^j) \tag{4}$$

式（4）和传统的需求方程不同，它并不是将需求写成一个关于第 i 个企业的证券价格的函数，而是写成了第 i 个企业的证券的单位美元回报的函数。不过，我们可以将式（4）改写成更“传统”的形式。借助莫迪格利安尼—米勒定理，我们可以不失一般性地假设：企业仅仅发行股票，因此其资产全部价值等同于企业的全部价值。因此，投资于第 i 个企业的单位美元回报也就等于：将第 i 个企业的全部回报除以其资产的全部价值。

$$e_i(\theta) = X_i(\theta)/V_i, \quad r = X_s/V_s \tag{5}$$

其中，$X_i(\theta)$ 是在状态 θ 下第 i 个企业的全部回报。下标 s 表示无风险行业。[18]将式（5）代入式（4），我们得到

$$V_i^j = D_i^{j*}(V_1, \cdots, V_n, V_s; X_1, \cdots, X_n, X_s, W_0^j) \tag{5'}$$

最后，将所有的个人加总，我们得到市场的“需求”方程：

$$V_i^d \equiv \sum V_i^j = \sum D_i^{j*} =$$

$$D_i^*(V_1, \cdots, V_n, V_s; X_1, \cdots, X_n, X_s; W_0^1, \cdots, W_0^m) \tag{6}$$

这样，我们就刻画了市场的“需求”方面。现在我们将对企业进行更详细的描述。

2. 企业行为

每一个企业都由一种生产技术来刻画，它决定了企业在状态 θ 下的产出：如果企业投资 I_i，并且选择生产计划 ξ_i，则：

$$X_i(\theta) = X_i(\theta, I_i, \xi_i) \tag{7}$$

企业存在多种生产计划，例如，对农场而言，比如说是否使用化肥，这在下雨的天气里能够增加产出（给定投入水平），而在干旱的天气里则会减产。

企业选择一定的投资水平和生产计划以最大化其原始股东的股票市场价

---

18 在本文中，我们通篇假设存在完全无风险的企业。如果没有这样的企业，只要假设存在一种无风险证券（没有拖欠风险的债券），这一分析就可以很轻易地被修正。这种情况下，市场分配仍然不是最优的。我们作出存在完全无风险的企业的假设，是因为我们集中关注的问题之一是：对于无风险行业和风险行业之间的资源分配，是否存在系统性的错误分配。

值，$V_i - I_i$。[19]难点在于确定风险企业投资水平或生产计划的变化对其股市价值的影响。我们将企业认知到的市场价值与其投资水平和生产计划之间的关系称为企业的估价函数（valuation function）。如果存在一个阿罗—德布鲁价格的完全集合——因此，$q(\theta)$ 表示下一期如果状态为 θ 则发放一美元这一承诺在当期的价格——V_i 的计算公式可以由此直接得出：

$$V_i = \sum_{\theta} X_i(\theta) q(\theta) \tag{8}$$

如果没有一个阿罗—德布鲁市场的完全集合，但是自然有多少种状态，就有多少个独立的企业，则企业就可以推断出在每一种自然状态下的收入的市场估价，也就是说，如果企业知道其他 V 个企业的市场估价（V 是自然状态的数量），以及它们在每种自然状态下的产出品，则企业可以找到一组 $q(\theta)$，使得

$$V_i = \sum_{\theta=1}^{v} q(\theta) X_i(\theta) \qquad i = 1, \cdots, v \leqslant n$$

用向量表示就是：[20]

$$q(\theta) = V \cdot X^{-1} \tag{9}$$

请注意，在这种情形下，企业必须知道其他企业在每种状态下的产出以及它们的市场估价。

但是，正如我们在引言中所提及的，经济中不仅没有阿罗—德布鲁证券的完全集合，而且企业的数量也少于自然状态的数量。

在下文将要被详细讨论的例子中，存在另外一个很自然的估价函数。在现阶段，我们关注的是理解模型的基本结构。为此，我们仅需要给出企业的一些估价函数，以此确定投资水平 I_i（I_i 是其他企业的产出和市场估价的函数，或者是一些市场决定的参数（如阿罗—德布鲁模型中的$q(\theta)$

19　如我们在引言中注意到的那样，当证券的数量少于自然状态的数量的时候，很明显，价值最大化并不是立足于股东的利益的。此处我们采用这个假设，是为了使我们的模型结果尽可能地和传统的阿罗—德布鲁模型的结果具有可比性。并且由于这是一条被广泛认定为企业会使用的行为准则，我们相信使用它的推论会非常有意义。参见 I. Fisher, *The Theory of Interest* (New York: Macmillan Co., 1930)；M. Jensen and J. Long，前引文献；F. Modigliani and M. Miller，前引文献；以及 J. E. Stiglitz, "Value Maximization," 前引文献。

20　假设"企业间相互独立"保证了 X^{-1} 的存在。关于市场证券和阿罗—德布鲁证券之间的关系的更详细的讨论请参见：D. Cass and J. E. Stiglitz, "The Structure of Preferences and Returns and Separability in Portfolio Allocation," 前引文献，以及 D. Cass and J. E. Stiglitz, "Risk Aversion and Wealth Effects on Portfolios with Many Assets," *Review of Economic Studies*.

的函数)。[21]

3. 市场均衡

最后，我们要讨论经济中均衡的条件。均衡是由以下这一系列变量描述的：所有企业的估价 V_i，投资水平 I_i，生产计划的选择 ξ_i。通过这些变量，使得下列条件得到满足：给定其他企业的投资水平、企业估价以及生产计划的选择，每个企业都相信自己正在最大化其原始股东的股市价值；投资需求的总和等于要素的总供给，即：

$$\sum_i I_i = \sum_j I^j \tag{10}$$

对不同证券（假定在各个自然状态下企业单位美元的回报是给定的）的总需求等于证券的总供给，即：

$$V_i = V_i^d \equiv \sum V_i^j \tag{11}$$

回顾一下，我们关注于第一期中 $n+2$ 个市场（$n+1$ 个证券市场和一个要素市场）中的均衡的决定因素。在第二期中没有经济行为；个人在第一期通过购买各种证券而“签订了一定的合约”，他们在第二期根据“合约”获得一定的商品并且将其消费掉。在第一期期初，给定要素的供给，个人对不同企业具有要求权（claim），他们相互之间进行此类要求权的交易。企业通过对下期产出的要求权进行交换得到所需的要素。它们确定投资水平和生产计划的选择，使得原始股东权益的价值最大化，其中的估价是由某个（暂未给出的）估价函数决定的。均衡要求企业的投资等于个人所供给的 I（要素市场的均衡）以及“交换均衡”（exchange equilibrium）——在给定的证券市场价格下，没有人愿意以某个企业的要求权交换另一企业的要求权。经济的时间进程和符合定义在表 1 中给出。

还需给出的是企业的估价函数的精确设定。我们将在一个具体的例子——均值方差模型——中给出这一函数。这一模型在不确定性和投资组合分析的文献中发挥了中心作用。

21 我们需要对企业的估价函数限定另一个条件；在均衡中，企业在投资水平和所采用的生产计划下所“预测”的价格必须正是其真正实现的价值。

表 1 两期经济

	第一期					第二期
	期初			期末		收入
	禀赋		禀赋的市场估价	投资组合配置（Max $Eu(Y^j(\theta))$）		（状态 $\theta = Y^j(\theta)$ 时的消费）
个人	要素所有权 对第 i 个企业的所有权（股份） 总的财富	I^j β_i^j	I^j $\dfrac{\beta_i^j(V_i - I_i)}{W_0^j}$	购买第 i 个风险企业的证券（所拥有的第 i 个企业的股份） 剩余资金投资于无风险资产	V_i^j $\gamma_i^j = \dfrac{V_i^j}{V_i}$ $W_0^j - \sum_{i=1}^{n} V_i^j$	$V_i^j e_i$ $(\gamma_i^j X_i(\theta))$ $r(W_0^j - \sum V_i^j)$
企业	最大化原始股东的股市价值 投资水平决策 生产计划决策 企业的市场估价			$\max(V_i - I_i)$ I_i ξ_i V_i		产出 $X_i(\theta) = X_i(\theta, I_i, \xi_i)$ 单位美元的回报 $\dfrac{X_i(\theta)}{V_i} = e_i(\theta)$
市场均衡			$\sum I_i = \sum I^j \quad V_i = \sum V_i^j$			

三、均值方差模型中投资水平的决定

1. 个人行为

我们假设个人通过衡量其均值 EY 和方差 σ_{Y^2} 来评估不同的收入模式；如果以 u（Y）标识以收入形式表示的效用方程，则

$$Eu(Y) = U(EY, \sigma_Y) \tag{12}$$

其中

$$\sigma_{Y^2} = E(Y - EY)^2 \tag{13}$$

这一假设的合理性和局限性在其他文献中已经得到了详细的探讨。[22] 为集中讨论我们所关注的问题，我们需要假设所有个人对每个企业在各种状态下的产出都有正确的（进而相同的）预期。另外，为了便于比较市场配置和最优配置，我们需要假设所有的个人都是相同的。[23]

众所周知，如果个人的期望效用可以写成收入的均值和标准差的函数，我们可以将投资组合配置问题分解为两个独立的问题：（a）找到投资于无风险资产的财富比例；（b）找到投资于风险资产的财富比例。[24] 事实上我们可以证明，[25] 所有的个人（无论其效用方程是怎样的，初始禀赋如何……）都会购买同如下比例一样的风险资产

---

22 特别是，我们要求个人有一个二次效用函数，并且所有回报的联合概率服从多元正态分布。关于这些条件的充分性的讨论，请参见：J. Tobin，"Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies*, vol. 25（Feb. 1958），pp. 65 – 86，以及"The Theory of Portfolio Selection," in Hahn and Brechling, eds.，*The Theory of Interest Rates*（London：Macmillan，1965）pp. 3-51. 关于这些条件的必要性的讨论，请参见：M. Rothschild and J. E. Stiglitz，"Increasing Risk：A Definition," *Journal of Economic Theory*, vol. 2（Sept. 1970），pp. 225-243。

23 然而，这一假设可以很轻松地去掉。P. Diamond，前引文献。

24 这正是托宾（Tobin 前引文献）证明的投资组合分离定理（portfolio separation theorem），要了解有关的完整讨论和对此理论的一般化，请参见 D. Cass and J. E. Stiglitz，"The Structure of Preference and Returns and Separability in Portfolio Allocation," 前引文献。

25 令 $\lambda = (\lambda_1, \ldots, \lambda_n)$，$\mu = (\mu_1, \ldots, \mu_n)$。则 $EY \equiv W_0 \, (r + \lambda \cdot (\mu - r))$，并且 $\sigma_Y = \lambda\sigma\lambda' W_0^2$。一个有效的投资组合必须在任何给定的均值下使方差最小化；构建拉格朗日函数

$$L \equiv \lambda\sigma\lambda' W_0^2 + q' \, (M - (\lambda \cdot (\mu - r) + r) \, W_0) \tag{13}$$

我们得到下列必要条件：

$$\lambda\sigma = q \, (\mu - r) \tag{14}$$

其中 $q = q'/W_0$，或者

$$\lambda = q \, (\mu - r) \, \sigma^{-1} \tag{14'}$$

式（15）可随即推导出。

$$\frac{\lambda_i}{\lambda_j} = \frac{\sum_k (\mu_k - r)\sigma'_{ki}}{\sum_k (\mu_k - r)\sigma'_{kj}} \tag{15}$$

其中，$e_i(\theta)$ 表示在状态 θ 下，投资于第 i 个风险资产的单位美元回报，$i=1, \cdots, n$；

$\mu_i = Ee_i$ 表示投资于第 i 个风险资产的单位美元平均回报；

$\sigma_{ij} = E(e_i - \mu_i)(e_j - \mu_j)$，表示投资于第 i 个企业和第 j 个企业的单位美元回报的协方差；

σ_{ii}表示投资于第 i 个证券的单位美元回报的方差；

σ'_{ij}表示 $\{\sigma_{ij}\}$ 的逆的第 ij 个因子的余因子；

λ_i 表示对第 i 个风险资产的投资占全部资产的比例；

r 表示投资于无风险资产（无风险企业）的单位美元回报。

我们以 σ_R 和 μ_R 分别表示投资于最优风险资产组合的单位美元回报的标准差和均值；令 λ_i^* 表示对第 i 个风险资产的投资占总风险投资的最优比例。

$$\mu_R \equiv \sum \lambda_i^* \mu_i \tag{16}$$

$$\sigma_R = \sum_i \sum_j \lambda_i^* \sigma_{ij} \lambda_j^* \tag{17}$$

剩下的资产组合配置问题就是将多大比例（以 a 表示）的财富分配到风险资产上。我们通过选择 a 来最大化以下公式：

$$EU(W_0(a\mu_R + (1-a)r), a\sigma_R W_0) \equiv EU(EY, \sigma_Y)$$

由此得到：

$$-\frac{U_1}{U_2} = \frac{\sigma_R}{\mu_R - r} \tag{18}$$

均值和标准差之间的边际替代率一定等于“调整”后的协方差（以无风险资产作为原点）。资产组合配置可以以我们所熟悉的均值方差图表来表示（图1），其中阴影部分表示以不同比例仅仅投资于风险资产可能获得的所有均值和标准差组合，R 给出了最优风险资产组合，S 表示全部投资于无风险资产所获得的最终财富，而 SR 表示可能性曲线。E 是均衡点。

2. 市场估价的均衡

使用均值方差模型的最大好处在于，在均衡的时候，各个企业的价值会以一种简单而符合直觉的方式彼此相关联：对于每一企业，我们可以计算出其不确定收入的确定性等价；其市场估价就是回报的确定性等价减去无风险资产的回报。其收入的确定性等价就是平均回报减去风险折扣因子和企业收

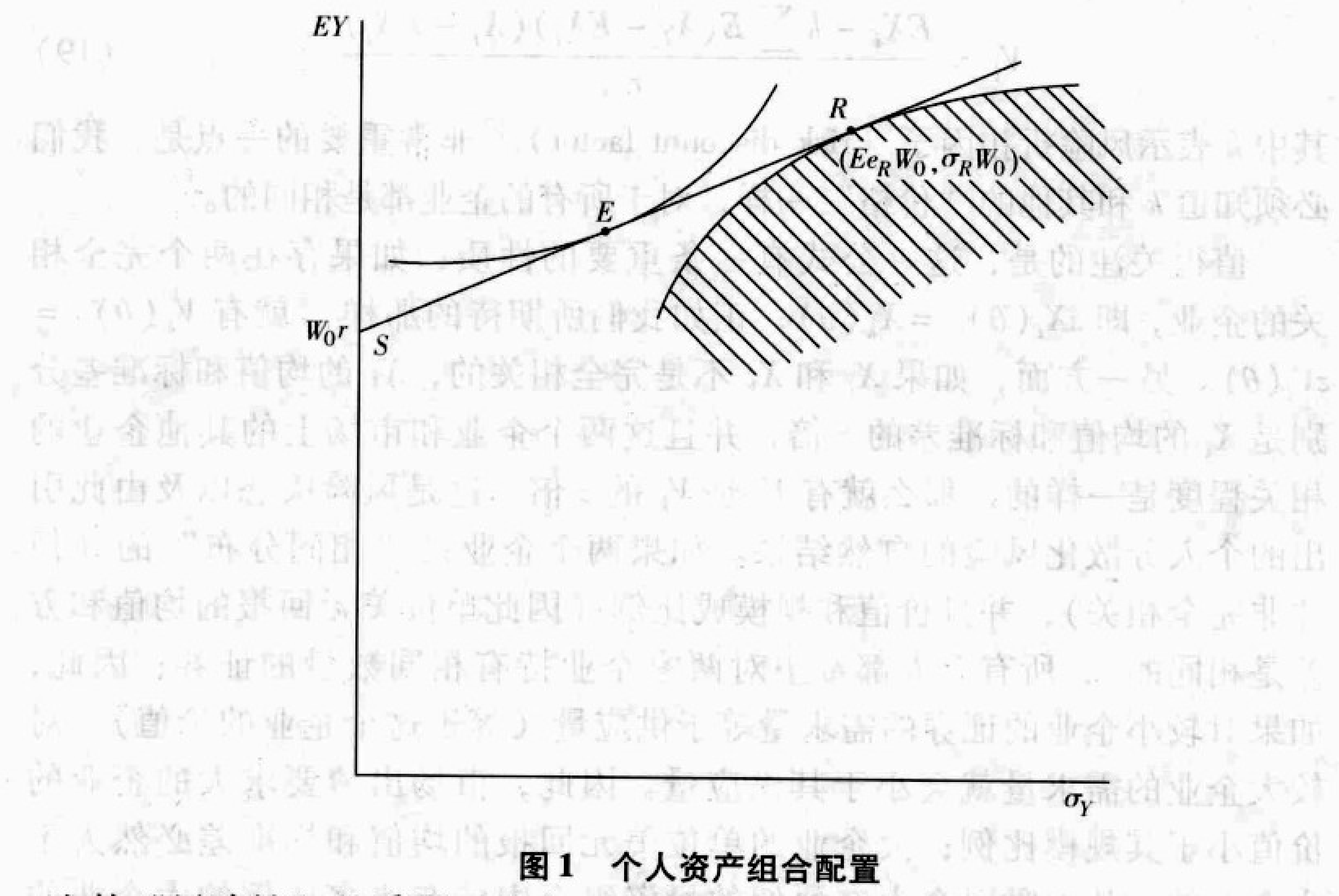


图 1　个人资产组合配置

入的"风险性"的乘积。其中，"风险性"就是其自身的方差和它与其他企业的协方差之和。由此，[26]我们得到：

26　这一结果最初由 J. 林特纳（J. Lintner）在下述文献中获得："The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, vol. 47（Feb. 1965）。其他相关文献包括：W. F. Sharpe, "Capital Assets Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, vol. 19（Sept. 1964）, pp. 425－442. J. Mossin, "Equilibrium in a Capital Assets Market," *Econometrica*, vol. 34（Oct. 1966）, pp. 768－783. 以下结果通过将式（5）代入式（14）获得，即

$$\sum_k E\lambda_k^j\left(\frac{X_i}{V_i}-\frac{EX_i}{V_i}\right)\left(\frac{X_k}{V_k}-\frac{EX_k}{V_k}\right)=q^j\left(\frac{EX_i}{V_i}-r\right)$$

将上式左右两边乘以 V_i 得到

$$\frac{1}{W_0^j q^j}\sum_k \frac{\lambda_k^j W_0^j}{V_k}(X_i-EX_i)(X_k-EX_k)=EX_i-rV_i$$

对所有 j 求和，并且知道在均衡的时候，

$$\sum_j \lambda_k^j W_0^j = V_k$$

我们可以得到

$$\frac{1}{\sum W_0^j q^j}\sum (X_i-EX_i)(X_k-EX_k)=EX_i-rV_i$$

由此，我们可以立刻得到式（19）。

$$V_i = \frac{EX_i - k\sum E(X_i - EX_i)(X_j - EX_j)}{r} \tag{19}$$

其中 k 表示风险折扣因子（risk discount factor）。[27] 非常重要的一点是，我们必须知道 k 和其他的“价格”一样，对于所有的企业都是相同的。

值得关注的是，这一公式有一条重要的性质：如果存在两个完全相关的企业，即 $zX_i(\theta) = X_k(\theta)$，正如我们所期待的那样，[28] 就有 $V_k(\theta) = zV_i(\theta)$。另一方面，如果 X_i 和 X_k 不是完全相关的，X_k 的均值和标准差分别是 X_i 的均值和标准差的 z 倍，并且这两个企业和市场上的其他企业的相关程度是一样的，那么就有 V_i 是 V_k 的 z 倍。这是风险厌恶以及由此引出的个人分散化风险的自然结果。如果两个企业是“相同分布”的（但并非完全相关），并且价值和规模成比例（因此单位美元回报的均值和方差是相同的），所有个人都希望对两家企业持有相同数量的证券；因此，如果对较小企业的证券的需求量等于供应量（等于这个企业的价值），对较大企业的需求量就会小于其供应量。因此，市场出清要求大的企业的价值小于其规模比例；大企业的单位美元回报的均值和标准差必然大于小企业的，从而吸引个人在他们的投资组合中持有更多比例的大企业的股份。

比例因子 k 主要取决于向量 I^j 和 β_i^j——经济中个人的数量和偏好，以及不同企业的产出模式 $X_i(\theta)$。但是，在特殊的两种情况下，k 的形式非常简单：

（a）如果所有个人都有一个二次效用函数，即

$$u^j(Y^j) = -e^{-a^jY^j}$$

27　将式（19）对所有的 i 加总，我们可以得到如下结果：

$$k = \frac{\sum EX_i - r\sum V_i}{\sum\limits_i \sum\limits_j E(X_i - EX_i)(X_j - EX_j)}$$

28
$$V_i = \{EX_i - k[\sum_{j\neq i,k} E(X_i - EX_i)(X_j - EX_j) + E(X_i - EX_i)^2(1+z)]\}/r$$

$$V_k = \{zEX_i - k[\sum_{j\neq i,k} z(E(X_i - EX_i)(X_j - EX_j) + E(X_i - EX_i)^2(z^2+z)]\}/r = zV_i$$

则[29]

$$k = \frac{1}{\sum \frac{1}{c^j} - \sum EX_i} \tag{20}$$

（b）如果所有个人的绝对风险厌恶系数为一个常数，即：

$$u^j(Y^j) = -e^{-a^jY^j}$$

并且如果所有的 x_i 都是联合正态分布的，从而 Y^j 是正态分布的，均值为 EY，方差为 σ_{Y^2}，则

$$U^j(EY,\sigma_Y) = Eu^j = -\exp\{-a^jEY^j + a^{j2}\sigma_{Y^2}/2\}$$

由此，[30]

$$k = 1/\sum(1/a^j) \tag{20'}$$

请注意，在这两种情况下，所有可获得的证券的方差并不影响 k 的大小；但是增加分摊风险的个人的数量会减少 k。[31]

---

29 证明如下：我们可以写成

$$Y^j(\theta) = (\sum \lambda_i^j(e_i - r) + r)W_0^j = \sum \gamma_i^j(X_i - rV_i) + rW_0^j$$

其中，$\gamma_i^j = \lambda_i^j W_0^j/V_i = V_i^j/V_i$ 等于由第 j 个人拥有的第 i 个企业的股份（在其最优化分配其投资组合之后）。因此关于 $\lambda_i^j(\gamma_i^j)$ 的最优选择就意味着

$$\frac{\partial EU^j}{\partial \gamma_i^j} = E\{(X_i - rV_i) - c^j(\sum_k (X_i - rV_i)(X_k - rV_k)\gamma_k^j + W_0^j r(X_i - rV_i))\} = 0$$

替换 $X_j - rV_j = X_j - EX_j + EX_j - rV_j$，对所有 j 加总，我们可以得到

$$\left(\sum\left(\frac{1}{c^j} - rW_0^j\right)\right)(EX_i - rV_i) = \sum_k E(X_i - EX_i)(X_k - EX_k) + (EX_i - rV_i)\sum_k (EX_k - rV_k)$$

给定 $\sum V_k = W_0^j$，可以得到

$$\frac{\partial \ln Eu^j}{\partial \gamma_i^j} = a^j(EX_i - rV_i) - a^{j2}\sum_k E(X_i - EX_i)(X_k - EX_k)\gamma_k^j = 0。$$

30 由下式得到：

$$(EX_i - rV_i)(\sum \frac{1}{c^j} - \sum EX_k) = \sum E(X_i - EX_i)(X_k - EX_k)$$

对所有的 j 加总，得到

$$\left(\sum \frac{1}{a^j}\right)(EX_i - rV_i) = E(X_i - EX_i)(X_k - EX_k)$$

这里 k 采用如此简单的形式是因为，在这个特例中对风险资产的需求并不依赖于财富水平。可参见如下文献：Cass－Stiglitz，"The Structure of Preferences and Returns and Separability"，前引文献。

31 P. A. 萨缪尔森（P. A. Samuelson）在下述文献中曾经观察到类似的结果："Risk and Uncertainty：A Fallacy of Large Numbers，" *Scientia*（April－May 1963），6th Series，57th year，pp. 1－6.

在后文中，将 k 表述为个人在均值和标准差之间的边际替代率是非常有用的。因为所有的个人以相同比例持有风险资产，我们可以将所有的风险性企业加总起来。令 V_R 表示加总的风险资产组合的价值，则

$$V_R = \sum_{i=1}^{n} V_i, X_R = \sum_i X_i, \mu_R = EX_R/V_R, \sigma_{R^2} = E(X_R - EX_R)^2/V_R^2$$

由式（18）得到

$$\frac{EX_R/V_R - r}{\sigma_R} = -\frac{U_2}{U_1}$$

即

$$V_R = \frac{EX + (U_2/U_1)(E(X_R - EX_R)^2)^{1/2}}{r} \tag{21}$$

将式（19）和式（21）进行比较，我们立刻得到

$$k = \frac{-U_2}{U_1(E(X_R - EX_R)^2)^{1/2}} \tag{22}$$

k 等于均值和标准差的边际替代率除以风险资产的标准差，同样，它是均值和方差的边际替代率的两倍。[32]

3. 企业行为

在前一小节中，给定 $X_i(\theta)$（给定他们的投资水平和生产计划决策），我们得到了不同企业的相对价值。现在我们转而讨论企业决策的决定因素这一重要问题。我们用一个具体的例子来展开讨论。在这个例子中，企业唯一的决策是 I_i。进而，我们假设所有的企业都是独立的，即：

$$E(X_i - EX_i)(X_j - EX_j) = 0, i \neq j \tag{23}$$

因此，如果 $\sigma X_i^2 = E(X_i - EX_i)^2$，均衡中企业的价值就仅仅取决于 EX_i 和 σX_i，以及市场参数 r 和 k。

$$V_i = \frac{EX_i - k\sigma_{x_i^2}}{r} \tag{24}$$

为描述第 i 个企业的技术，我们仅需要对均值和方差是如何取决于投资水平的进行具体的描述：

32 必须弄清楚的是，虽然我们最初采用了均值和标准差来表示效用函数，但这只是一个主要基于分析方便考虑的武断的选择，我们可以简单地写成：

$$Eu = v(EY, \sigma_{Y2})$$

并且通过 v，而不是 μ 来进行分析。

$$EX_i(I_i) = h_i(I_i) \quad (25a)$$

$$\sigma_{X_i}(I_i) = g_i(I_i) \quad (25b)$$

平均回报随着投资水平的增加而增加，但是增加的速度越来越慢。

$$h'_i > 0, \quad h''_i \leqslant 0, \quad (26a)$$

同时，标准差是一个随着 I_i 的增加而增加的凸函数，

$$g'_i > 0, \quad g''_i \geqslant 0 \quad (26b)$$

（我们所需要的仅仅是 $g''_i g_i + g'^2_i \geqslant 0$。对于无风险行业而言，$g_s \equiv 0$）

企业通过选择其投资水平来最大化其原始股东的股市价值，即如果 $V_i(I_i)$ 表示当投资水平为 I_i 时企业的价值，则

$$\frac{d(V_i - I_i)}{dI_i} = 0$$

或者

$$\frac{dV_i}{dI_i} = 1 \quad (27)$$

对于无风险行业而言，这就意味着如果它采取竞争性行为，并且给定利率水平，则由此可得

$$V_s = h_s(I_s)/r$$

从而

$$h'(I_s) = r \quad (28a)$$

无风险资产的投资回报率恰好等于对无风险行业投资的边际生产率。这一结论和实际应得的结果相同。

对于风险性企业而言，计算其他备选投资水平的结果只是稍稍复杂一点。企业可以计算出在其他备选投资水平 $\chi_i(I_i)$ 下的随机产出的确定性等价，其价值就是：

$$V_i \equiv \frac{\chi_i(I_i)}{r}$$

因此，

$$\chi'_i = r$$

确定性等价加号就是均值减去一个“风险溢价”。其中，“风险溢价”就是方差乘以一个“风险折扣因子”。也就是说，我们可以考虑企业在出售两种产品：均值和方差，前者的“价格”是 $1/r$，后者的“价格”是 $-k/r$（即方差的增加会降低企业的价值）。如果给定这些价格（k 和 r），企业的股市价值最大化就意味着（根据式（24）和式（27））

$$\frac{h'_i(I_i) - 2kg'(I_i)g_i}{r} = 1 \tag{28b}$$

根据式（26）这一假设，式（28）可以转换成

$$I_i = I_i(k,r) \tag{29}$$

其中

$$\frac{\partial I_i}{\partial k} = \frac{2g'_i g_i}{h''_i - k(g''_i g_i + g'^2_i)} < 0 \tag{30a}$$

并且

$$\frac{\partial I_i}{\partial r} = \frac{1}{h''_i - k(g''_i g_i + g'^2_i)} < 0 \tag{30b}$$

4. 市场均衡

现在，我们可以很轻松地将均值方差模型的全部竞争市场均衡描述出来了。式（28）定义了在每一个 r 和 k 价值下的投资水平。对于每一个可以出清投资市场的 r 和 k 的价值，即

$$I_d(r,k) \equiv \sum_i I_i(r,k) = \sum I_j \equiv I \tag{10'}$$

式（19）定义了一对可以出清证券市场的 r 和 k 的价值。可以同时满足式（10′）和式（19）的 r 和 k，以及与之相连的 I_i 和 V_i 构成了一个竞争均衡。

如果 m 个消费者都是相同的，并且风险性行业中的所有企业都是相同的（但是独立分布）——非常清楚的是，这些风险性企业的投资水平也会是相同的，那么这个例子中的市场均衡就可以用图表描述出来。

首先，我们来描述均值方差的可能性曲线。如果 I_R 是投资于风险性行业的投资总量，在第二期单位美元的回报的均值和方差由下式给出：

$$EY = \{h_s(I - I_R) + h_R(I_R)\}/m \tag{31a}$$

$$\sigma_{Y^2} \equiv \frac{g_{R^2}(I_R)}{m^2} \equiv \frac{n}{m^2}g_i^2(I_R/n) \tag{31b}$$

其中，

$$h_R(I_R) \equiv nh_i(I_R/n)$$

在式（26）描述的技术假定下，在相关定义域内——均值和方差随着对风险性行业的投资分配的增加而增加的区域，即

$$\frac{dEY}{d\sigma_Y} = \frac{h'_R - h'_s}{g'_R} > 0 \tag{32}$$

可能性曲线是凹的,[33]如图 2 所示。

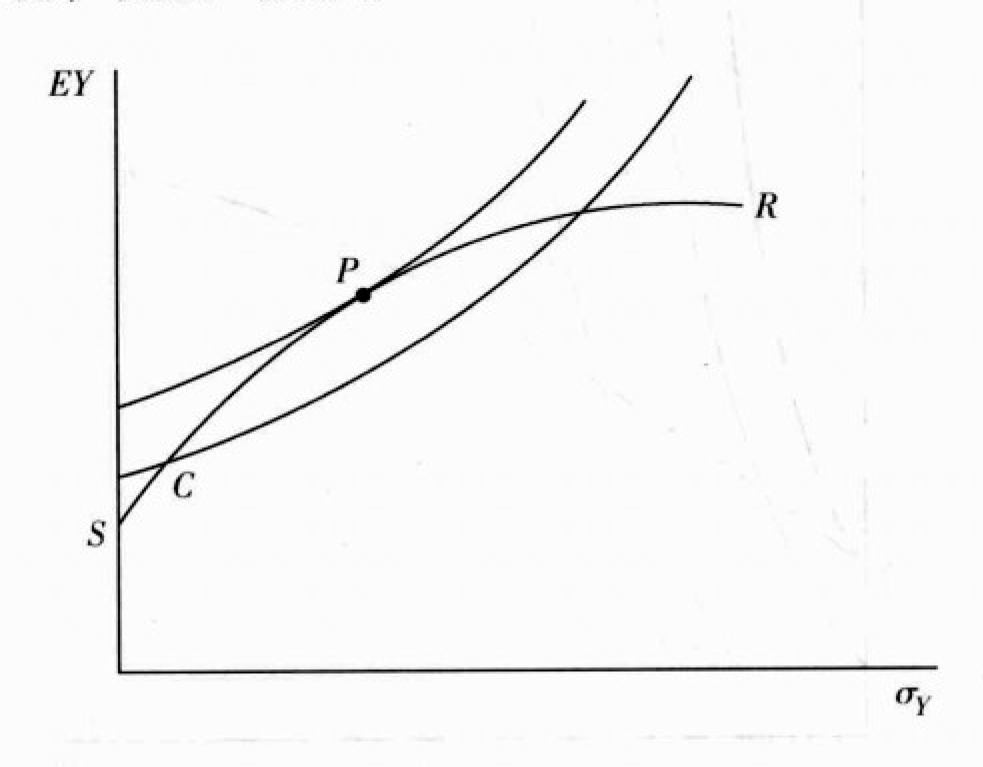


图 2 投资的市场配置和最优配置

通过增加 r 减少 k，我们可以得到可能性曲线上的所有的点（使用式（10′）和式（30））。由此我们可以计算出，在给定点上无差异曲线的斜率需要是多大，才能诱使个人按照证券供给量来持有证券。利用式（22），我们可以得到

$$-\frac{U_2}{U_1} = kg_R(I_R) \tag{33}$$

我们需要对式（33）和式（32）进行比较。将式（28）代入式（32），我们可以得到

$$\frac{dEY}{d\sigma_Y} = 2kg_R(I_R) \tag{34}$$

因此，可能性曲线的斜率总是恰好是可以使给定点落在可能性曲线上的无差异曲线的斜率的两倍。C 点给出了本例中的竞争均衡。

到目前为止，我们还不能保证得到一个内点均衡（interior equilibrium）；我们很有可能得到一个仅限于无风险行业或者仅限于风险性行业的市场均衡，如图 3a 和图 3b 所示。从现在开始，我们将集中关注内点均衡。

对于前文介绍的两个特例——具有二次的效用函数以及具有常绝对风险厌恶系数的情况，我们可以很轻松地解出均衡。在后一个特例中，式（20′）

---

33 $\frac{d^2EY}{d\sigma_{Y2}} = \left(\frac{h''_R + h''_s}{g'_R} - \frac{dE\,Y\,g''_R}{d\sigma_Y\,g'_R}\right)\frac{m}{g'_R} < 0$

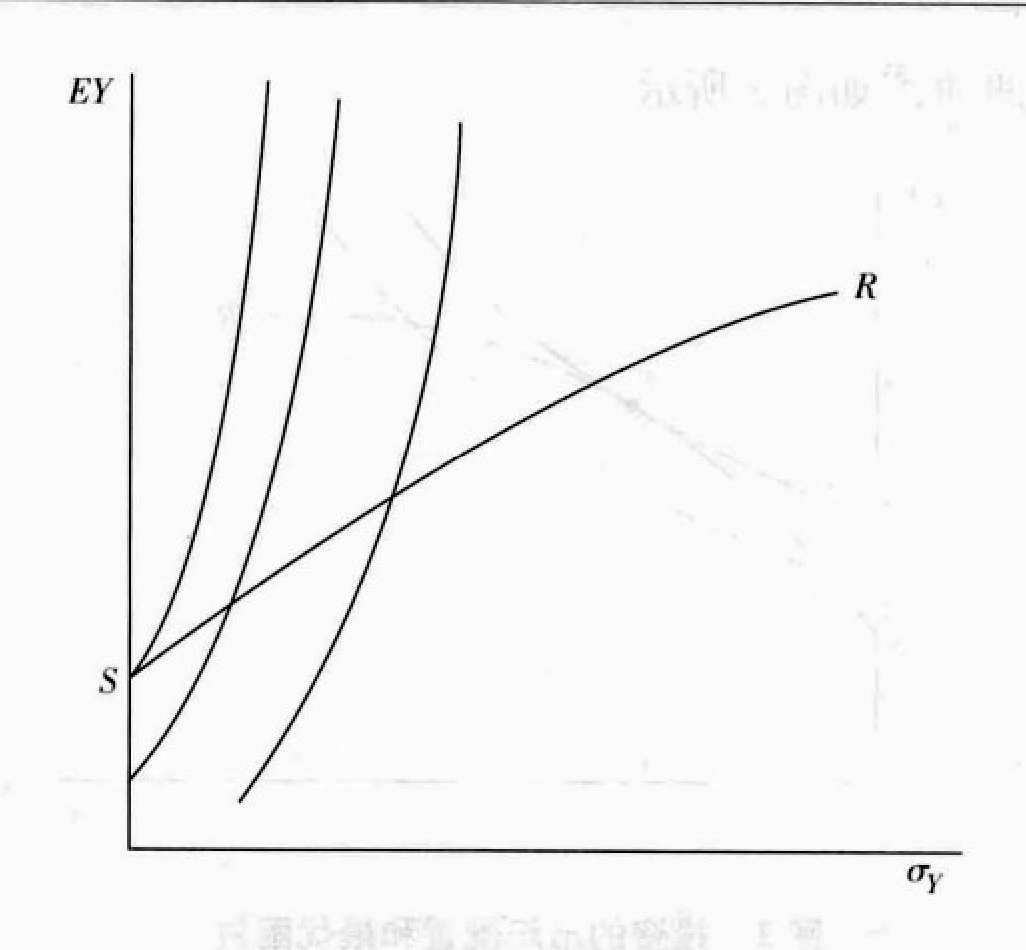


图 3a　集中于风险性行业

表示，市场均衡要求 $k = 1/\sum(1/a^j) \equiv k^*$。因此，我们很轻松地找到了满足下式的 r 的值

$$I_d(r,k^*) = I \tag{35}$$

如果

$$I(0,k) = \infty \tag{36a}$$

$$I(\infty,k) = 0 \tag{36b}$$

我们总是可以在 $0<r<\infty$ 的范围内找到满足式（35）的唯一的 r 值。[34] 如果 $h'_s(0) > r^* > h'_s(I)$，则均衡是内点解。

同样，对于效用函数是二次型的情况，k 必须满足式（20）。为了通过图表解出构成均衡的 r 和 k 的值，我们首先观察到投资市场的出清方程确定了 r 和 k 的一条斜率为负的曲线；（在假设式（36b）下）这一曲线与横轴交于 $r\equiv r_{max}$（r_{max} 为有限值）；在大于 r_{max} 的区域，即使风险的价格为零，投资的需求也会少于供给；当 r 变小时，（在假设式（36a）下）曲线趋近于垂直线 $r\equiv r_{min}$。在小于 r_{min} 的区域，即使风险的价格为无穷，也会存在对投资品的过度需求（见图4）。沿着投资品均衡曲线，我们可以看出 $\sum EX_i$ 是 r 的增函

34　如果对于任意 i，$\lim_{I_i\to 0} h'_i(I_i)\to\infty$，并且对于所有 i，$\lim_{I_i\to\infty} h'_i(I_i)\to 0$，这就显然满足条件。

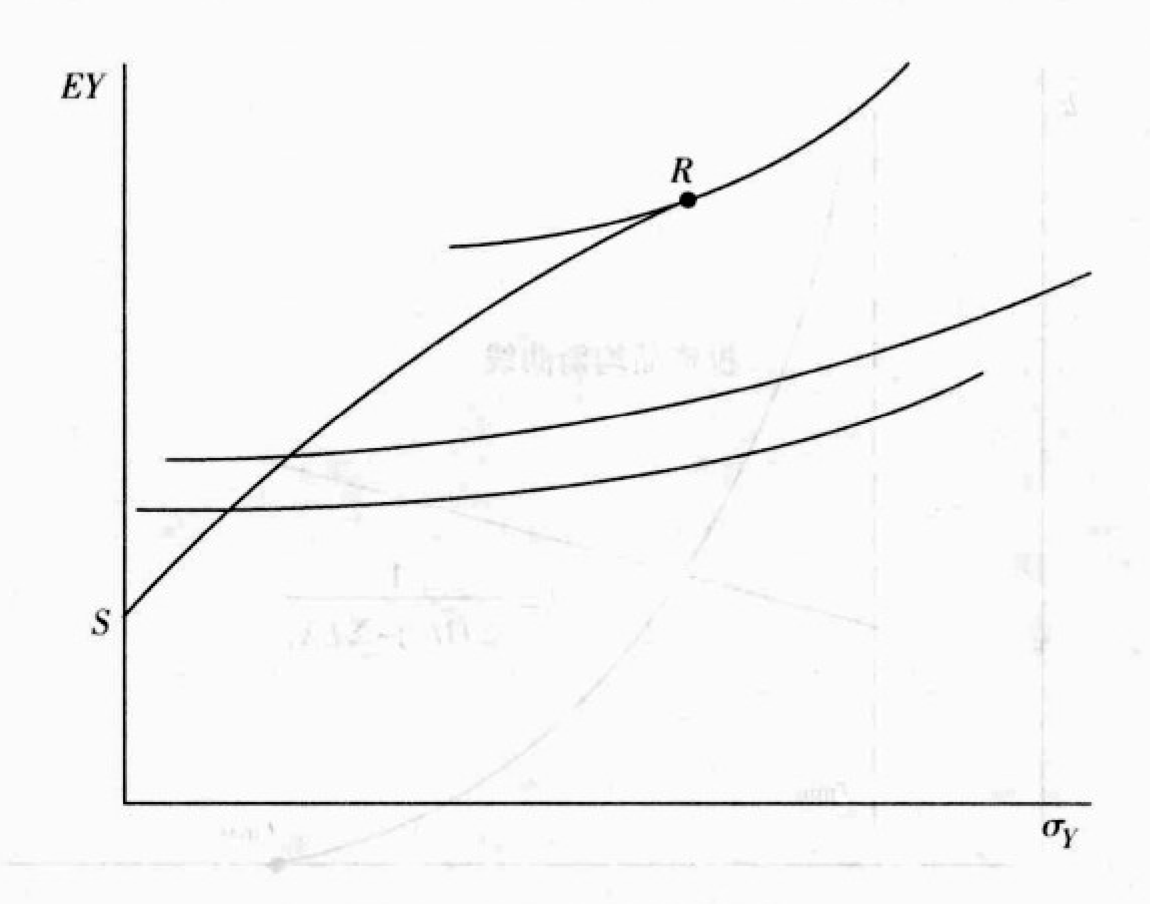


图 3b　集中于无风险行业

数；因此，结合式（14），k 是 r 的增函数。[35] 显然，两条曲线必定相交，而且仅相交一次。

5. 帕累托最优

现在我们希望将第 4 点描述的市场均衡和最优均衡进行一下比较。我们要最大化 $U(EY, \sigma_Y)$，其中 σ_Y 由式（31）定义。最优化要求

$$h'_i - h'_s = -\frac{U_2}{U_1}\frac{g_i g'_i}{g_R} = -\frac{U_2}{U_1}\cdot\frac{1}{2g_R}\cdot\frac{dg_i^2}{dI_i} \tag{37a}$$

对第 i 个资产的平均投资回报的边际增加和对无风险资产的平均投资回报之间存在差异，这一差异和方差的边际增加之间成一定比例关系，同时比例因子就是边际替代率除以风险资产的标准差的两倍。这需要和市场均衡进行一定的比较。运用式（22）和式（28），可以把市场均衡表述成：

35　在式（10）定义的曲线上，我们要求

$$\sum \frac{dI_i}{dr} = \sum \frac{\partial I_i}{\partial r} + \frac{dk}{dr}\sum \frac{\partial I_i}{\partial k} = 0$$

但是，由式（30），

$$\frac{dI_i}{dr} = \frac{-1 + \frac{dk}{dr}\frac{h'_i - r}{k}}{h''_i - k(g''_i g_i + g'^2_i)}$$

由式（26）可知，分母都是负的，而分子是 h' 的单调函数，因此那些有着较大 h' 的企业，$dI_i/dr > 0$，对于有着较小 h' 的企业，$dI_i/dr < 0$，但是 $dEX_i/dr = \sum h'_i dI_i/dr$ 并且结果相当直接。

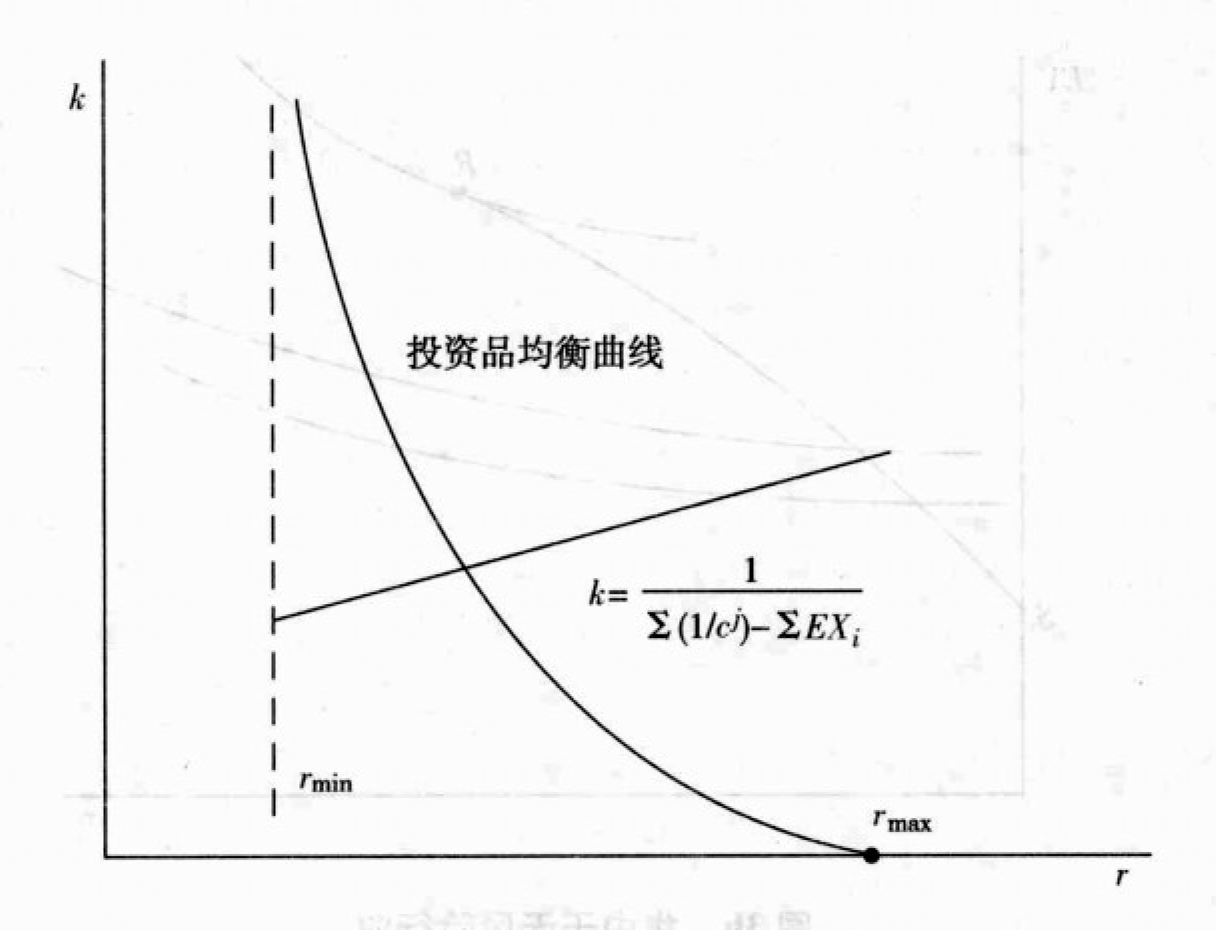


图4　均衡的决定因素（k, r）（二次效用方程）

$$h'_i - h'_s = -\frac{U_2}{U_1}\frac{2g'_i g_i}{g_R} = -\frac{U_2}{U_1}\cdot\frac{1}{g_R}\cdot\frac{dg_i^2}{dI_i} \qquad (37b)$$

这些方程都是相同的，除了比例因子是最优配置的两倍，即风险的“价格”是它应有价格的两倍。

这对资源配置的影响可以通过图解的方式表现出来，如下文所述：

假设所有的风险性企业都是相同的。我们可以回忆到，在上一小节中讲到可能性曲线是凹的。因此最优点 P 就是无差异曲线和可能性曲线的切点（图2）。比较 P 和 C，可以很清楚地发现 P 总是在 C 的右边：较之最优配置，市场配置导致在风险性行业的投资较少。

更宽泛地来说，我们很有可能证明，尽管市场均衡总是在社会可能性曲线上，[36]即是“有效的”，它往往导致对无风险行业的过多投资。

6. 非独立回报

然而，当我们放松独立的假设的时候，经济就不再是“有效”的了，即，它将在均值方差可能性曲线“之下”运行。为了证明这一点，假设对任一企业的回报可以用如下方程表示：

$$X_i(\theta, I_i) = h_i(I_i) + g_i(I_i)\epsilon_i(\theta) + m_i(I_i)M(\theta) \qquad (38)$$

36　有效性要求在 $\sum h_i$ 为常数的约束下，最小化 $\sum g_i^2$，即 $vg'_i g_i = h'_i - h'_s$。其中 v 是一个比例因子。但是市场上的投资方程（28）有着相同的参数形式，因此市场配置是有效的。

其中

$$E\epsilon_i = EM = 0 \tag{38a}$$

因此，

$$EX_i = h_i \tag{38b}$$

另外，[37]

$$E\epsilon_i\epsilon_j = 0, E\epsilon_i M(\theta) = 0 \tag{38c}$$

$$EM(\theta)^2 = 1, E\epsilon_i^2 = 1 \tag{38d}$$

因此，

$$E(X_i - EX_i)^2 = g_i^2 + m_i^2 \tag{38e}$$

$M(\theta)$ 是所有企业共有的市场因子（商业周期的状态）。式（38）表明去除了公共市场因子后的残差是在企业之间独立的。

因此第 i 个企业的市场估价就是（将式（38）代入式（19））

$$V_j = \frac{h_i - k(g_i^2 + m_i^2 + m_i \sum_{i \neq j} m_j)}{r} \tag{39}$$

一个企业的价值决定因素有三个：均值、方差以及企业回报的“商业周期”的相关性。

给定利率、风险折扣因子以及其他企业的投资（因此，由其他企业引起的“周期性波动的大小”(magnitude of the cyclical fluctuations) $\sum_{i \neq j} m_j$ 是给定的)，如果企业最大化其原始股东权益的价值，则

$$h'_i - r = k(2g_i g'_i + 2m_i m'_i + m'_i \sum_{i \neq j} m_j) \tag{40}$$

另一方面，均值方差有效边界（efficiency frontier）可以通过下式进行参数化定义[38]

$$h'_i - r = \kappa(g_i g'_i + m_i m'_i + m'_i \sum_{i \neq j} m_j) \tag{41}$$

比较式（40）和式（41），我们可以立刻得到如下这个结论：相对于协方差而言，市场配置对自方差给予的权重是其“应该”给予的权重（即沿

---

37 式（38d）中给出的条件是已经标准化了的。

38 我们想要在 $\sum g_i^2 + 2\sum m_i m_j = \sigma^2$ 约束下，最大化 $\sum h_i(I_i) + h_s(I - \sum I_i)$。构建拉格朗日方程

$$L = \sum h_i + h_s - \frac{\kappa}{2}(g_i^2 + 2\sum m_i m_j - \sigma^2)$$

并且对 I_i 求导，我们就可以得到式（41）。

着有效边界）的两倍；事实上，必须清楚的是，最优化要求企业不应对自方差比协方差更加看重——在均衡中，所有的证券都“集结”在一起成为一个多元基金。式（41）证实了这一点。

配置失当的原因在于企业之间的协方差起到了类似于外部性的作用。一个企业的投资水平的变化会影响到其他所有企业的价值；尽管它对任何单个企业的价值影响非常小，但当将对所有企业的影响加总之后，这一效应不容忽视。

这对投资配置失当的方向的意义是显而易见的。假设存在两个风险性行业。为了简化，令 $h_i = h'_i I_i$，$m_i = m'_i I_i$，$g_i = g'_i I_i$，其中 h'_i，m'_i，g'_i，是常数。假设这两个行业有同样的平均回报。然后，令单位美元投资回报的方差 $\sigma'^2_{xi} = m'^2_i + g'^2_i$，则

$$\left(\frac{I_i}{I_j}\right)_M - \left(\frac{I_i}{I_j}\right)_0 = \frac{m'_i m'_j(\sigma'^2_{xi} - \sigma'^2_{xj})}{(2\sigma'^2_{xi} - m'_i m'_j)(\sigma x'^2_i - m'_i m'_j)} \tag{42}$$

其中 $(I_i/I_j)_M$ 和 $(I_i/I_j)_0$ 分别表示市场配置中对行业 i 的投资和对行业 j 的投资的比例，以及最优配置中对行业 i 的投资和对行业 j 的投资的比例。那么，如果 $m_i m_j > 0$，即，这两项正相关，则市场对较低方差的行业分配的资源相对太少（若 $m'_j < m'_i + g'^2_i/m'_i$ 或者 $m'_j > m'_i + 2g'^2_i/m'_i$）。其他的情况留给读者讨论。

式（40）和式（41）有一个非常有趣而且重要的含义：我们观察到，如果自方差相对于协方差较小，由式（40）得到的投资水平和式（41）得到的投资水平非常近似；事实上，两者不仅是“近似”有效，而且是“近似”帕累托最优。[39]

四、技术选择

在前一节中，我们假设企业不能选择技术，考虑了投资的市场配置。与之同样重要的问题是，市场价值最大化的企业是否对技术选择做出“正确”的决策；也就是，如果企业可以降低它的方差——但是要以均值减少或者与

39 也就是说，帕累托最优要求

$$h'_i - r = \left(-\frac{U_2}{U_1}\right)\left(\frac{1}{\sum_i g_i^2 + \sum_i \sum_j m_i m_j}\right)^{1/2}(g_i g'_i + m'_i \sum m_j)$$

同时，由式（40）得到（运用式（22））

$$h'_i - r = \left(-\frac{U_2}{U_1}\right)\left(\frac{1}{\sum g_i^2 + \sum \sum m_i m_j}\right)^{1/2}\{(g_i g'_i + m'_i \sum m_j) + (g_i g'_i + m_i m'_i)\}$$

如果 $g_i g'_i + m_i m'_i / g_i g'_i + m'_i \sum m_j$ 很小，则两个方程近似相同。

市场的共线性增加为代价——企业是否可以做出正确的“得失”权衡(trade - off calculation)?

我们在这里得到的结果和在上一节得到的结果非常类似：如果所有的企业都是相互独立的，经济就在有效边界上运行；唯一的扭曲来自于均值和标准差之间的边际转换率（有效边界线的斜率）和边际替代率（无差异曲线的斜率）的不一致。另一方面，如果企业不是完全独立的，经济就不是在它的均值方差边界上，相对于协方差而言，它会对自方差给予太多的权重。

1. 相互独立的企业

在均值方差模型中，由于企业是相互独立的，市场股价仅依赖于产出品的均值和方差，我们可以通过以下函数充分刻画（相关的）技术集：

$$I_i = \Omega_i(\sigma_{x_i}, EX_i) \tag{43}$$

给定投资水平，我们可以通过减少均值来降低方差，同时，增加投资水平可以使我们获得较高的均值和标准差。参见图5a。

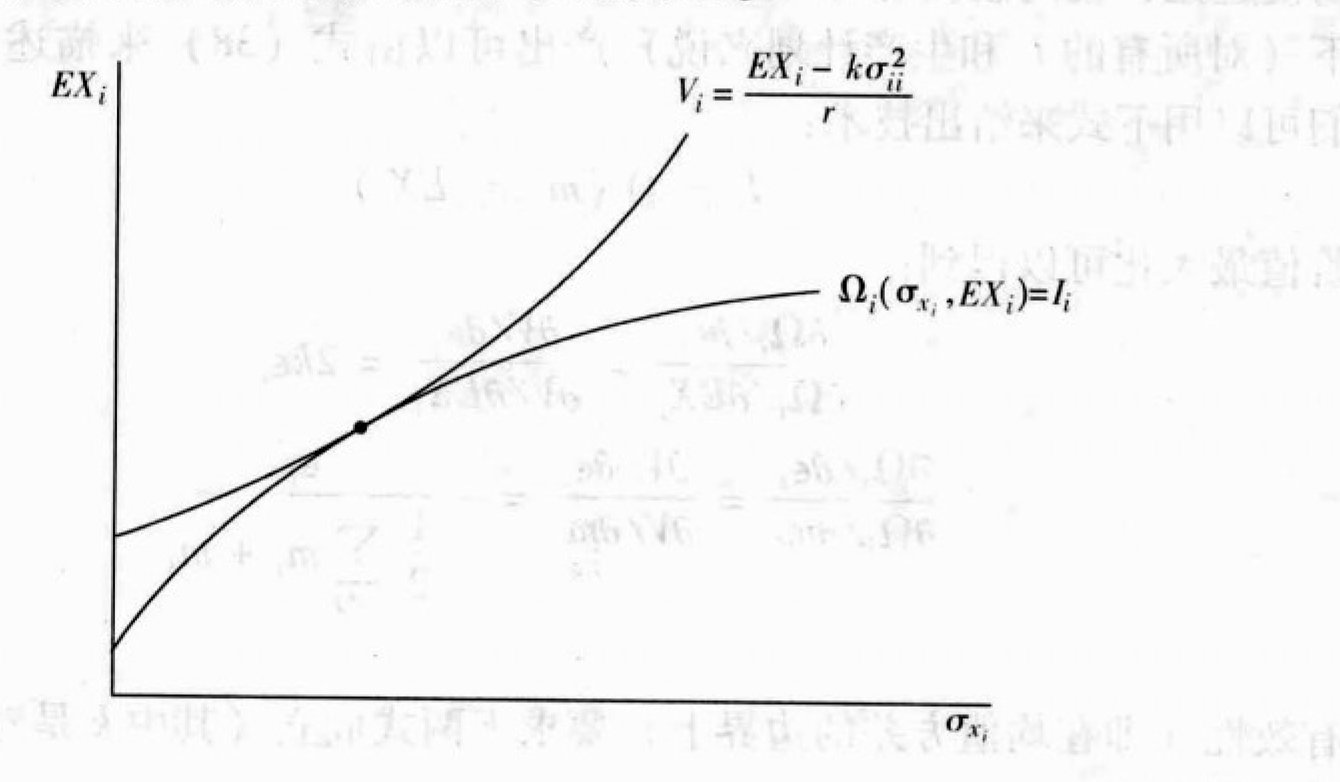


图5a　技术选择：独立的企业

给定 k 和 r，一个价值最大化的企业要选择 σ_{x_i} 和 EX_i，从而[40]

$$\frac{1}{r} = \frac{\partial \Omega_i}{\partial EX_i}$$

$$-\frac{k}{r} = \frac{1}{2\sigma_{x_i}} \frac{\partial \Omega_i}{\partial \sigma_{x_i}}$$

40　为保证有内点解，我们仅需下述条件：在给定 I_i 下，标准差的持续减少只能通过 EX_i 更大的持续地减少才能换得。

即

$$\frac{\partial EX_i}{\partial \sigma_{x_i}} = -2k\sigma_{x_i} \tag{44}$$

有效性要求

$$\min_{\text{s. t.}} \frac{\sum \sigma_{x_i}^2}{\sum EX_i} + h_s(I - \sum \Omega_i) = M$$

也就是说，如果 q 是关于上述约束的拉格朗日乘数，则，

$$2\sigma_{x_i} = qh'_s \frac{\partial \Omega_i}{\partial \sigma_{x_i}}, 1 = h'_s \frac{\partial \Omega_i}{\partial EX_i}$$

由此可知，竞争性经济是有效的，但并非是最优的。

2. 非相互独立的回报

当企业不是相互独立的时候，对技术的具体描述只是稍微复杂一点。为简便起见，我们假设对所有企业只有一个唯一的市场因子，在不同自然状态下（对所有的 I_i 和生产计划来说）产出可以由式（38）来描述。这样，我们可以用下式来给出技术：

$$I_i = \Omega_i(m_i, \epsilon_i, EX_i) \tag{45}$$

价值最大化可以得到：

$$\frac{\partial \Omega_i / \partial \epsilon_i}{\partial \Omega_i / \partial EX_i} = \frac{\partial V / \partial \epsilon_i}{\partial V / \partial EX_i} = 2k\epsilon_i \tag{46a}$$

$$\frac{\partial \Omega_i / \partial \epsilon_i}{\partial \Omega_i / \partial m_i} = \frac{\partial V / \partial \epsilon_i}{\partial V / \partial m_i} = -\frac{\epsilon_i}{\frac{1}{2}\sum_{i \neq j} m_j + m_i} \tag{46b}$$

有效性（即在均值方差的边界上）要求下两式成立（其中 k 是对所有 i 共同的比例因子）：

$$\frac{\partial \Omega_i / \partial \epsilon_i}{\partial \Omega_i / \partial EX_i} = 2\kappa\epsilon_i \tag{47a}$$

$$\frac{\partial \Omega_i / \partial \epsilon_i}{\partial \Omega_i / \partial Em_i} = \frac{\epsilon_i}{\sum_{i \neq j} m_j + m_i} \tag{47b}$$

比较式（46b）和式（47b），我们可以在此看到价值最大化的企业对于协方差部分（$\sum_{i \neq j} m_i$）给予太小的权重。这两个界的比较反映在图 5b 中（要提供自方差越低的产出品，困难就越大。同样，要提供和市场关联越小的产品，困难也同样越大。因此，在固定的 I_i 和 EX_i 下，我们假设 ϵ_i 是关于 m_i 的凸

函数)。

由这些结果可以得到一个直接的推论：如果经济中存在两个不同的项目，它们成本相同、均值相同，但其自方差以及与其他项目的协方差不同，总的市场方差跟由哪家企业从事这一项目无关；要对这两个项目做出选择，而且我们赋予不同的企业同样的选择机会，各个企业所做的选择通常都不一样。

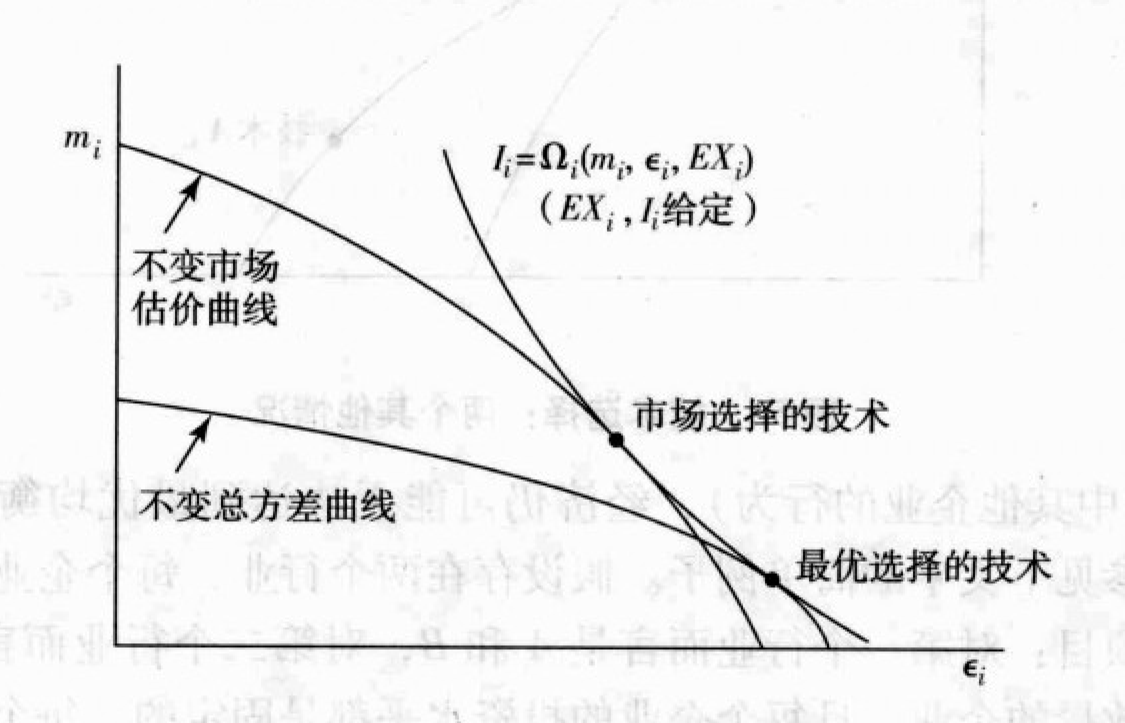


图 5b　技术选择：自方差对于市场的协方差

即使自方差项相对于协方差而言很小，企业在 ϵ_i 和 m_i 之间的“权衡”也是其应有值的两倍。当然，如果 ϵ_i 对所有 i 都很小，则这一结果就是可以忽略的。在 $\epsilon_i \equiv 0$ 的限制下（技术可以由 $I_i=\Omega_i$（m_i，EX_i）描述），如果存在大量的企业（m_i 远远小于 $\sum_{i\neq j} m_j$），则经济是“近似”有效和帕累托最优的。

3. 离散选择

这些结果相当令人满意。它们很符合我们的直觉：如果企业“出售”的唯一“风险”关于公共市场因子，并且个体企业相对于整个市场很小，那么我们可以得到通常的关于帕累托最优的竞争性结果。当企业同时“出售”一些和市场无关的风险时，它就对这些风险赋予了太大的权重。

为了得到这一结果，我们假设企业面临技术的一个“连续统”。在现实中，企业可能面临一系列离散选择；或者，如果存在一个技术的“连续统”，它可能要被定义在关于（ϵ_i，m_i，EX_i）的更小的取值范围内，这样才更有可能产生角点解。对自方差的关注可能导致对于生产计划的错误选择（参见图 5c）。

4. 结构性的无效率

更令人震惊的结果是，即使所有的企业都采取最大化社会福利的行为

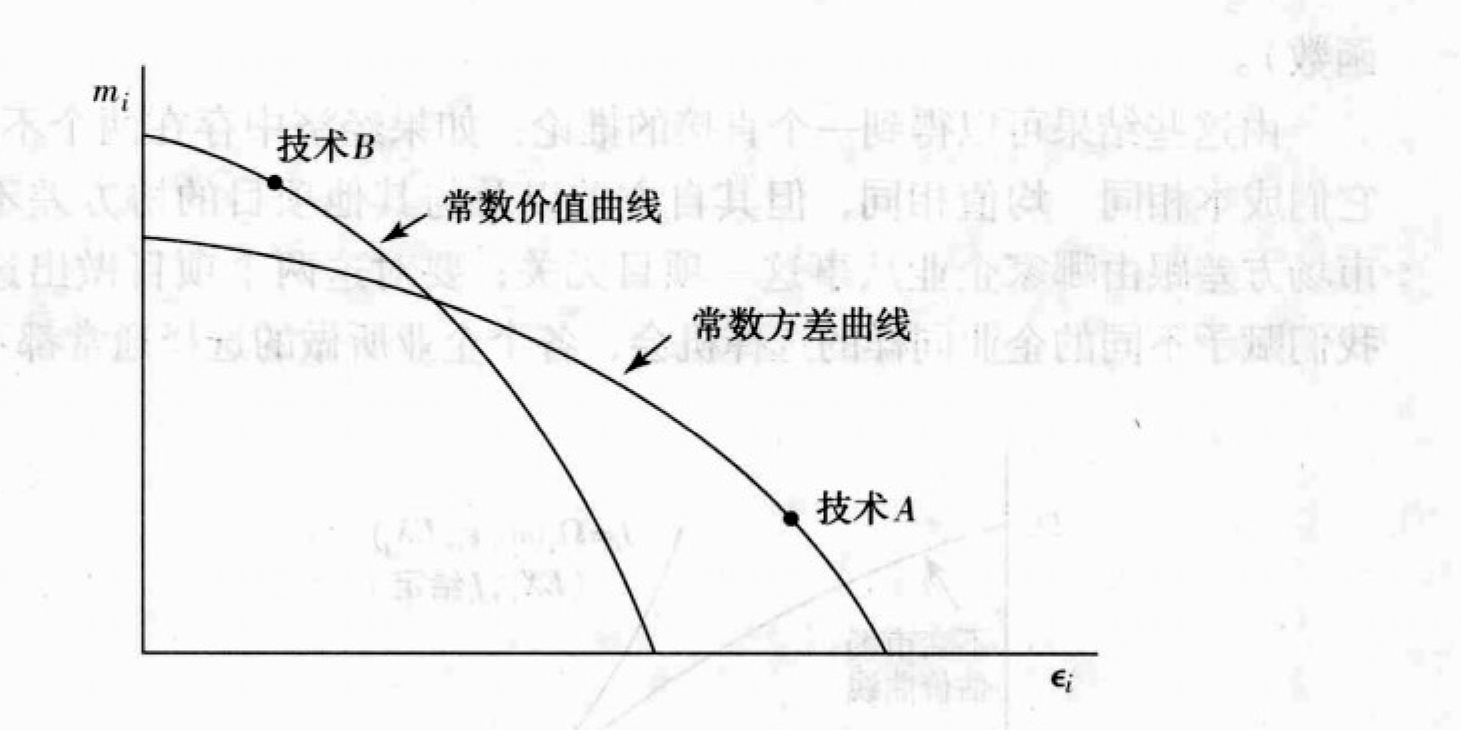


图 5c 技术选择：两个其他情况

（给定经济中其他企业的行为），经济仍可能无法达到最优均衡。要验证这一点，可参见下文中的简单例子。假设存在两个行业，每个企业都有两个备选的投资项目：对第一个行业而言是 A 和 B，对第二个行业而言是 C 和 D。给定一定数量的企业，且每个企业的投资水平都是固定的，每个企业面临的唯一问题就是在第一个行业用 A 还是 B，在第二个行业用 C 还是 D？对第一个行业而言，无论采用什么技术，企业的回报都是独立的。在第二个行业的情况也一样。所有项目的平均投资回报都是相同的，并且其自方差都是相同的。项目之间唯一的不同在于它们和其他行业中项目的协方差是不同的。令 C_{ij} 表示第一个行业中任一企业第 i 个项目的回报与第二个行业中任一企业的第 j 个项目的回报之间的协方差，我们对 C_{ij} 作如下假设：[41]

$$C_{AC} < C_{AD} \tag{48a}$$

$$C_{BC} > C_{AC} \tag{48b}$$

$$C_{BD} < C_{BC} \tag{48c}$$

$$C_{AD} > C_{BD} \tag{48d}$$

$$C_{AC} < C_{BD} \tag{48e}$$

给定其他企业的选择，假设每个企业选择能够最小化社会方差的项目（技术）（或者，我们可以令企业选择最大化自身价值的项目）。

在这个经济中，存在两个均衡。在其中一个均衡中，第一个行业里的所有企业都选择 A，第二个行业里的所有企业都选择 C；而在另一个均衡中，第一个

---

41 很容易就可以证明这些假设与方差协方差矩阵的正定性不一致。

行业里的所有企业都选择 B，第二个行业里的所有企业都选择 D。为证实这一点，我们观察到，由于我们假设了行业内部的独立性、相同的均值和自方差，唯一关键的是与另一行业的企业所采取的项目的协方差问题。如果第一个行业中的所有企业选择 A，那么第二个行业中的所有企业会选择 C，因为 $C_{AC} < C_{AD}$；相反，如果第二个行业中的所有企业都选择 C，那么第一个行业中的所有企业都会选择 A，因为 $C_{AC} < C_{BC}$。类似地，如果第一个行业中的所有企业都选择 B，那么第二个行业中的企业都会选择 D，因为 $C_{BD} < C_{AD}$。如果第二个行业中的所有企业都选择 D，那么第一个行业中的企业都会选择 B，因为 $C_{AD} > C_{BD}$。[42]

但是在均衡（AC）中，每个人的状况都比在均衡（BD）中好，因为 $C_{AC} < C_{BD}$。

可能有人会考虑兼并的情况，如果一个行业中的某个企业和另一个行业中的某个企业合并，由此导致的技术决策的协调会保证（AC）配置会被采用。可惜这是不对的。因为重要的互动发生在某一行业中的一个企业和另一行业的所有企业之间。总的来说，第二个行业中任何单个企业的行为通常不会影响第一个行业中企业的技术选择。为确保最优化，必须对整个经济中的项目（技术）选择进行协调；若存在不确定性，但不存在相机要求权（阿罗—德布鲁）市场的一个完全集合，那么我们就会在有效分权（efficient decentralization）问题上碰到严重的问题。

这是一个关于可能被称为经济中的结构性无效率。它无法通过任何单一生产者的边际改变而得到修正。

五、对各种关于市场均衡的模型的评论

在绪论的参考文献中提到，除了阿罗—德布鲁模型，戴梦德还精确构建了一个股市模型。[43]这个模型的中心结论是：投资配置是（受约束的）帕累托最优，与此处的结论截然相反。通过和汽车市场进行类比，我们更好地理解戴梦德的模型。汽车可以被看成一个复合商品；[44]如果有足够多不同种类

---

42 在这个问题中，我们假设企业只能在两个项目中二选一。但如果企业可以选择对两个项目的投资比例，也可以得到非常类似的结果，当然这需要在式（48）之外再给出一些限定。

43 前引文献。

44 正如下述文献中的讨论：K. J. Lancaster, "A New Approach to Consumer Theory," *Journal of Political Economy*, vol. 74 (April 1966), pp. 132 – 157; Z. Griliches, "Hedonic Price Indexes Revisited: Some Notes on the State of the Art," *Proceedings of the American Statistical Association*, 1967, Business and Economics Statistics Section, pp. 324 – 332.

的汽车，人们就可以推断尾板、烙钢、规格变化、乙烯基材料座位、布料座位等的市场估价。但是，如果给定生产商制作的汽车种类已经确定，这些“合成”价格就没有意义了。这时，生产商唯一关心的是其制造的复合商品的价格。如果生产商是价格接受者，我们可以很容易地看出市场均衡将是一个“受约束的”帕累托最优，即给定制造的汽车种类，生产的汽车数量是“正确的”。沿着莫迪格利安尼—米勒定理的思路，戴梦德认识到这也适用于风险资产。如果企业在不同自然状态下只有一种收入模式，那么我们只需要知道一单位该模式下收入的“价格”。事实上，我们可以完全避免明确涉及消费者或生产者方面的不确定性；消费者的预期效用就是他们所拥有的每个收入模式的“单位”（units）数（S_i）的函数——$EU=\phi$（S_1，…，S_n），同时，企业的市场估价是它们所供给的单位数的函数：$V_i=V_i(S_i)$。如果企业间相互“竞争”，即每一单位回报模式的价格是给定的，那么，很显然，产生的均衡是受约束的帕累托最优。

这一模型的局限是很明显的：（a）它没有对技术选择的基本问题给出解答；（b）只有在很紧的条件下——回报模式不随着企业投资水平的改变而改变（乘积性不确定性）——它才能对投资水平问题给出解答；事实上，运营规模（scale of operation）的变化常常导致带有不同回报模式的不同生产过程；（c）其所基于的行为学准则很值得商榷：戴梦德做出的如下假设是否合理？——当企业并非完全相关的时候，它们相信规模翻一番能够使其市场价值也翻一番。做出以下假设是否更加合理？——企业将认识到事实上更经常出现的情况是：企业规模翻一番，其市场价值将小于翻一番的价值。[45]

阿罗—德布鲁模型和戴梦德模型都试图建立一个和传统的没有确定性模型本质上相同的模型，来分析不确定性下的情况。我认为，这让我们对传统模型的缺陷有了更深入的认识，而非对经济有任何实质的帮助。

我们已经试图建立一个简单的替代的模型。前文已详细讨论的那个例子运用了众所周知的均值方差模型；这既是该模型的局限，也是其优势。因为它很有可能通过很简单的参数化来描述企业决策过程。它更大的长处在于：企业确实是价格的接受者，市场上的利率和风险折扣因子都已给定。事实

45　V. 史密斯（V. Smith）曾经强调，更困难的是，除了在本处分析的具有同质的期望（homogeneous expectation）的均值方差模型里，进行价值最大化的企业往往会违背无破产这一假设；企业的财务政策确实非常关键；由于建立的模型不能解释企业财务政策的决定因素，因此它们是不完全的。

上，我们建议企业采用的“价格”体系具有所有正常价格体系的性质；特别是，在均衡时，所有企业的价值可以用同样的估价方程算出（正如在阿罗—德布鲁模型中那样）。[46]在某些限定性的特例中，它还有其他的性质。这些特例中没有技术选择，并且相对于企业和经济中其他企业的协方差而言，其自方差是几乎可以忽略的，市场配置是近似帕累托最优的；在这些情况下，企业可以简单地忽略它们的自方差部分（它们和市场的其他部分不相关）。然而，因果实证主义和计量研究都认为：自方差对于解释市场估价有着重要作用；[47]同时认为企业对和商业周期无关的投资机会是“风险厌恶”的；[48]并且，尽管证券交易市场可能是现存的最具竞争性的市场之一，但企业在发行证券的时候并不是完全竞争者：它们意识到（至少在相对于模型的时间跨度而言的短期）规模增大引起的市场价值增长比例小于规模增大的比例。这些关于风险资产市场的特征正是我们在模型中试图获取的。

由于企业与其他企业的协方差大小不合适，并且企业认识到，在短期内它们的价值不能和规模同比例增长，因此，市场经济不是帕累托最优的[49]这一模型的结果清楚地证明了：股市并不像我们所认为的那样（尤其是，不像以前的股市模型得到的结果），是配置资源的一个“有效”机制。对这一模型的扩展（比如说允许企业自由进入）加强了此处报告的结果。在那种情况下，不仅风险性行业的总体投资水平太低，而且企业的数量也不是最优

46 在这一点上，其他的“价格”体系有时是失败的。例如，人们认为企业认为自身在出售“均值”和“标准差”。因此，如果企业的规模翻一番，则其价值也会翻一番；因此，市场配置将是帕累托最优的。但是，没有一个所有企业共有的标准差的“价格”可以正确地“预期”企业的价值；例如在独立的情况下，如果企业给每一单位标准差的“定价”是 $k\sigma_{x_i}$，它将正确地“解释”其市场价值，但是这个价格对于不同的企业是不同的。一个类似的建议是，企业认为它们并非在出售标准差，而是在出售“对市场财富总方差的贡献”。由于类似的原因，这也是错误的，我将在附录中给出更详细的证明。

47 请参见 G. W. Douglas，“Risk in the Equity Markets：An Empirical Appraisal of Market Efficiency，” *Yale Economic Essays*，vol. 9（Spring 1969），pp. 3 - 45 and J. Lintner，“Security Prices and Risk：The Theory and a Comparative Analysis of A. T. and T. and Leading Industrials，”这是一篇 1965 年 6 月 24 日在“管制公共效用的经济学”（Economics of Related Public Utilities）大会上报告的文章，由道格拉斯引用。事实上，这两个研究都认为，企业应当对其自方差赋予比本模型中更大的权重。同时参见 M. H. Miller 和 M. Scholes 的“Rates of Return in Relation to Risk：A Reexamination of Some Recent Findings，” mimeo，Chicago，1970。

48 如果自方差相对于和市场的协方差而言很小，企业可以轻易忽略其自方差；它们会通过市场多元化而消除掉。

49 除了在上文提及的限制性情况下。

的。即使企业是相互独立的，经济也会在均值方差边界之下运行（允许企业的数量变化）。

六、结论

竞争经济就好像是有“看不见的手”引导着，达到帕累托最优配置，长期以来，这一直被视为竞争经济最具吸引力的特征之一。在没有不确定性的静态经济中，这一结论的粗略证明及其反命题的证明——每一个帕累托最优配置都可以由竞争均衡获得（对初始禀赋进行一定的重新分配）——已经得到展示；同时，这一理论的局限，外部性的重要性，规模效应递增（非凸性）以及公共品问题已经得到了广泛的讨论。另外，阿罗和德布鲁已经证明了如何将这一模型*正式地*扩展到包含动态经济和不确定性的情况。但是这一模型所要求的市场并未出现在任何可观察到的经济中。事实上，我们已经考虑了一种情况，其中所有的个人都以同样的比例购买阿罗—德布鲁证券（只要证券可以获得）。至于消费者，他们对拥有由一个共同基金提供的证券非常满意。在这种情况下，如果与共同基金相比，在离散证券之间交易的成本哪怕有极小的增加，也很有可能只存在有限的这样的市场。有阿罗—德布鲁市场提供的，用以指导企业进行决策的价格对于企业做出正确的配置决策是必需的。

若存在股市，在分析投资资源的配置时，我们面临两个问题：

（1）不同的资本品（不同的证券）的价格是否正确地反映了这些商品（证券）的社会回报？或者是这些价格反映了我们所预期的别人对该投资品（证券）的预期价格？（凯恩斯[50]认为是这样的。）事实上，哈恩（Hahn）、谢尔（Shell）和斯蒂格利茨认为，若不存在期货市场的完全集合，不同资本资产的竞争性配置可能不是有效的（事实上，根据他们的分析，配置看起来不是有效的）。[51]

（2）就算不同证券的价格正确地反映了社会回报，我们是否可以得到一个有效的投资模式？这正是本文主要关注的问题。通过考虑一个两期模型，并在大多数分析中假设每个个人对每一企业的回报模式的认识都一致，我们可以从第一个问题中脱身而出。当然，我们已经论证了，投资在风险资

50 J. M. Keynes, *The General Theory* (New York: Harcourt, Brace, and World, 1936).

51 F. H. Hahn, "Equilibrium Dynamics with Heterogeneous Capital Goods," *The Quarterly Journal of Economics*, vol. 80 (Nov. 1966), pp. 633 – 646; K. Shell and J. E. Stiglitz，前引文献。

产和无风险资产之间的配置。在各种风险资产之间的配置，以及技术的选择不会是最优。

这些结论表明，竞争这一“看不见的手”可能并不像我们原先想的那么有效。在某种意义上，竞争性市场的这些局限性比原先所注意到的要严重得多。因为我们在本文中关注的是资源在各种投资机会间的配置——这是所有经济活动中最基本的活动之一。另外，外部性可以通过适当的税收政策来修正，而公共物品必定要政府提供；如果仅有少数几个行业存在规模报酬递增，那么这些行业可被国有化，或者被政府管制，而其他经济组织仍为私有；各种投资机会间竞争性配置失当问题看起来不容易被修正。我们所提出的竞争性经济的一些问题同时也是分权化的社会主义经济的问题。集权化决策机制能否在实际中给出一个更优的配置，这还未有定论。

附录

法玛（Fama）[52]以及其他一些人提出了另一种股价函数：企业考虑其对期望市场财富和市场财富的总方差的贡献，即，

$$V_i = \left(\frac{1}{r}\frac{\partial EX}{\partial I_i} - \frac{\kappa}{r}\frac{\partial \sigma_x}{\partial I_i}\right) I_i \tag{A.1}$$

其中，EX 是总的（所有企业的）平均收入，σ_x^2 是总方差，$\kappa = U_2/U_1$ 是均值和标准差之间的边际替代率。法玛的结论中提到，“所有这一切有着重要的含义：在双参数模型（two - parameter model）中，不同企业的股份是完全替代品；从企业在资本市场上出售的物品来看，每个企业都不是唯一的”。

能使这些令人愉快的论断有着一定有效性的一个特殊例子是：技术是规模报酬不变的（投资翻倍可以使各种自然状态下的产出翻倍），[53]并且估价函数正是戴梦德以及莫迪格利安尼—米勒所使用的股价函数。

为获得这些结论，令 g_i^2 是每单位投资的自方差，h_i 是每单位投资的平均回报，m_i 是与其他所有企业（每单位投资）的协方差，则

$$\frac{\partial \sigma_x}{\partial I_i} = (I_i g_i^2 + m_i)/\sigma_x, \frac{\partial EX}{\partial I_i} = h_i \tag{A.2}$$

因此，将式（19）重写，我们可以得到一个与（A.1）形式相同的公式：

52　E. Fama, “Risk, Return, and Equilibrium,” U. of Chicago, mimeo, 1968.

53　分析很容易被推广至不确定性增加的更一般情形。

$$V_i = I_i\left\{h_i - \kappa \frac{\partial\sigma_x}{\partial I_i}\right\}/r$$

其中（结合式（22）），

$$\kappa = k\sigma_x = -U_2/U_1 \qquad (A.3)$$

假设 κ 和 $\partial\sigma_x/\partial I_i$ 独立于企业的投资水平，这相当于假设企业价值和投资成比例。因此，在此情形下，这一估价函数导出帕累托最优性的原因与法玛的“双参数模型”无关，与不同证券是相近替代品这一假设也无关；相反，它直接由技术的特殊性质以及企业的价格接受行为产生，无须对效用函数的性质作任何限定就可以保证其正确性。更宽泛地讲，企业的价值无法由式（A.1）的均值来表示（并且，如果可以由此表示的话，就无法确保由此导致的配置是最优的）。为证明这一点，我们可以考虑下面这个简单的例子，假设第 i 个企业的回报独立于其他企业的回报，并且

$$EX_i = h_i I_i$$

$$\sigma_i = g_i I_i^v$$

那么

$$\frac{\partial\sigma_x}{\partial I_i} = \frac{vI_i^{2v-1}g_i^2}{\sigma_x}, \frac{\partial EX_i}{\partial I_i} = h_i$$

以及

$$V_i = I_i\frac{(h_i - kg_i^2 I_i^{2v-1})}{r} = \frac{\left(h_i - \kappa\frac{1}{v}\frac{\partial\sigma_x}{\partial I_i}\right)I_i}{r}$$

仅当 $v=1$ 时，即随机的规模报酬不变，此时才与式（A.1）的形式相同。

这一法则还存在其他的问题：

（i）假设 $\partial\sigma_x/\partial I_i$ 独立于 I_i 是不合理的；考虑一个极端的情况：所有的企业都独立分布，且 $v=1$。

无论企业有多少，$\partial(\partial\sigma_x/\partial I_i)/\partial I_i$ 和 $\partial\ln(\partial\sigma_x/\partial I_i)/\partial\ln I_i$ 都不为零：

$$\frac{\partial(\partial\sigma_x/\partial I_i)}{\partial I_i} = \frac{g_i^2}{\sigma_x}\left(1 - \frac{I_i^2 g_i^2}{\sigma_x^2}\right) \to \frac{g_i^2}{\sigma_x}$$

假如随着企业数量的增加，由单个企业提供的方差比例就会趋近于0，类似地

$$\frac{\partial(\ln\partial\sigma_x/\partial I_i)}{\partial\ln I_i} = 1 - \frac{I_i^2 g_i^2}{\sigma_x^2} \to 1$$

（ii）另外，企业如何估计它对“风险”的贡献从未得到详细讨

论——除了随机规模报酬不变这一特例。这并不是一个容易观察到的市场参数。[54]

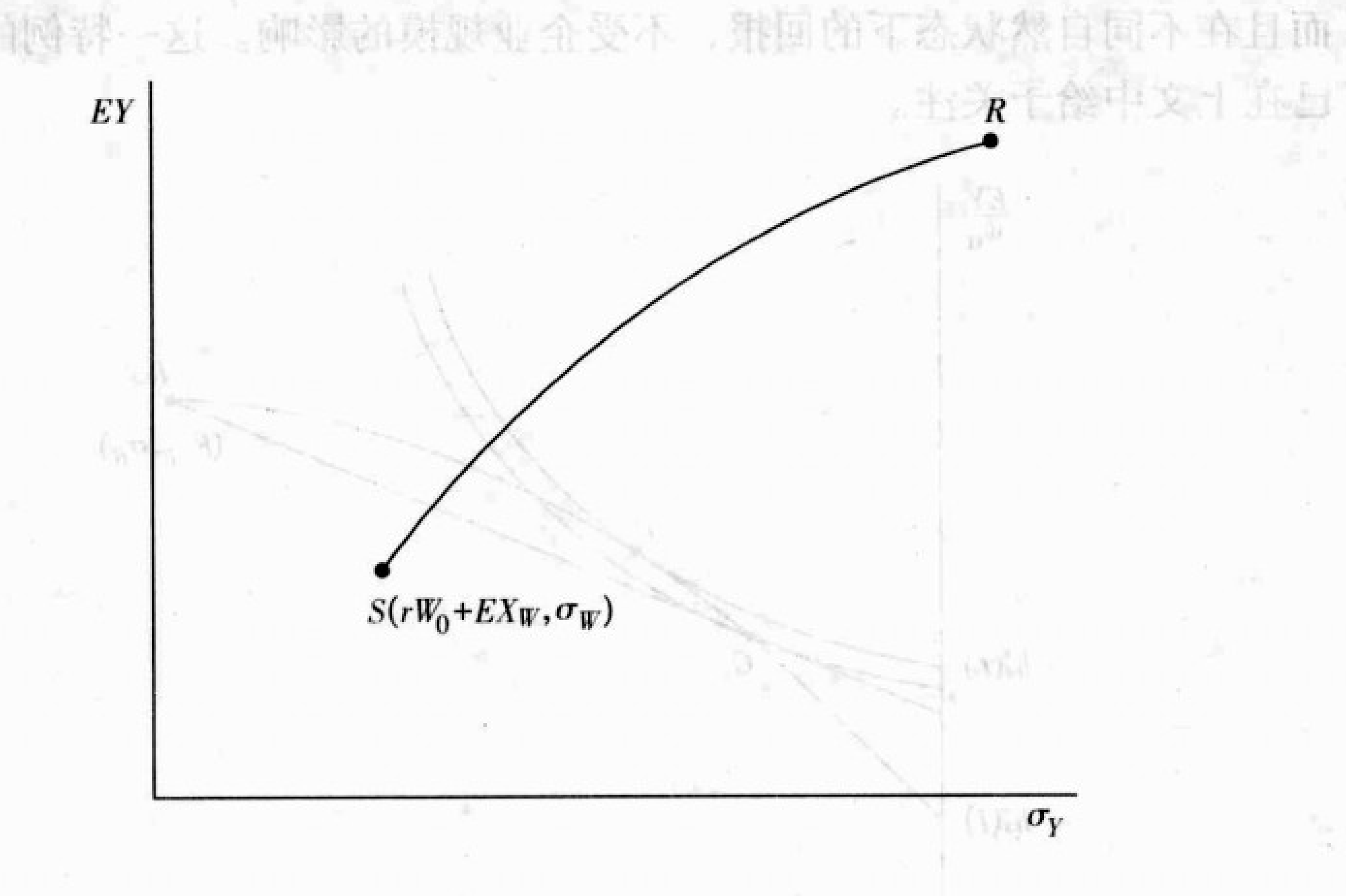


图6　可变工资收入时的机会曲线

简而言之，尽管那些信奉市场的人（在令人难以理解的基础上）认为，在不存在不确定性的情况下，投资的市场配置是最优的，他们很欢迎这一估

54　以上方法还有一个缺陷，即它无法扩展到下述情况：所有个人都有风险性工资收入，其均值为 EX_w^i，标准差为 σ_w^i。为简化起见，我们假设所有的证券都独立于工资收入。可能性曲线将不再是一条直线；相反，它是一个有着正斜率的凸函数（如图6所示），这一点很容易证明。这意味着不同个人的均值标准差边际替代率是不同的；不再存在前文讨论的均值标准差之间交换的单一市场。然而，我们可以证明，式（19）那样的估价函数仍然是有效的。

许多关于投资组合分析的讨论都通过回报率的均值和标准差来展开。由于个人最终感兴趣的是消费，似乎通过他期末提供的财富（消费）来评价各种投资项目的方法更加合理。然而，需要指出的是，如果通过回报率的均值和方差来衡量其他收入流，即便在风险性行业存在随机规模报酬不变，且企业作为价格接受者，市场配置也不会是帕累托最优的。为证明这一点，假设无风险行业具有一个规模报酬递减的生产函数，并且存在一个风险性行业与之竞争。社会可能性曲线将是一个凹函数，如图7所示，其与纵轴的交点为 $h's(I)$。给定利率，个人将面临一条直的可能性曲线。如果对无风险行业分配的资源是 I_s，个人的可能性曲线与纵轴的交点是 $h'_s(I_s)$。若 $I_s<I$，则 $h'_s(I_s)>h'_s(I)$。由于假设在风险性行业中是随机规模报酬不变，并且风险性行业中的企业假设有单位投资的价值率（the ratio of the value per unit of investment）与投资水平无关。如只投资于风险资产，在竞争均衡中均值和标准差的边际替代率。但是请注意，这里的最优配置在风险性行业中的投资水平又是比市场配置中的水平要高。

价函数，但实际上，这一估价函数并不怎么吸引人；这一函数能够良好定义并且“起作用”的特例已在第五部分详细讨论，其中的企业不能选择技术，而且在不同自然状态下的回报，不受企业规模的影响。这一特例的局限性早已在上文中给予关注。

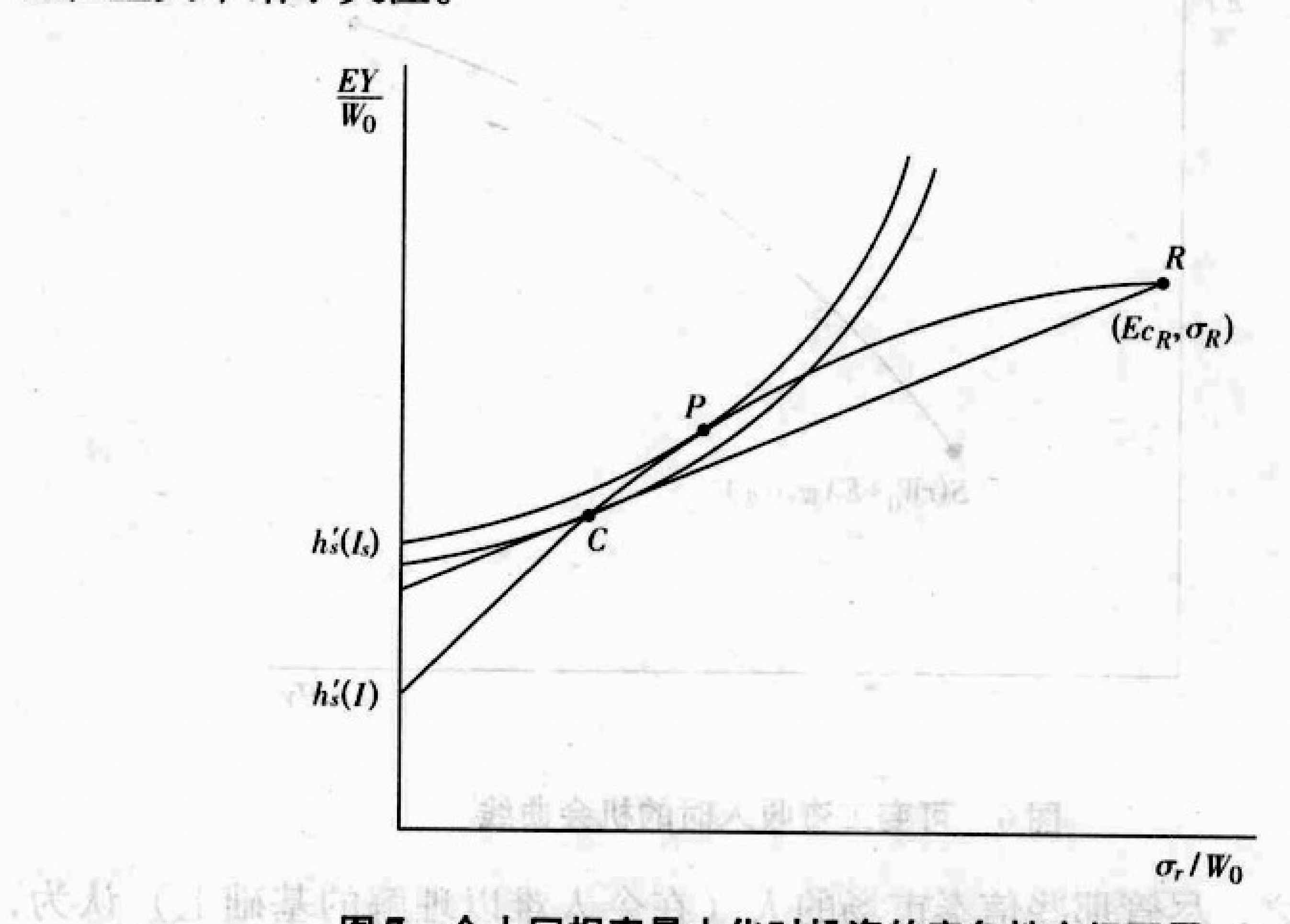


图 7　个人回报率最大化时投资的竞争性市场配置

价值最大化与企业的可选择目标*

近期关于企业行为的文献有两种截然不同的分析框架。一方面，一部分文献试图进一步扩展传统的竞争性企业利润最大化模型，将其从我们熟悉的静态模型扩展至动态模型，并引入不确定性。这一分析框架认为，企业的目标是让股市价值最大化，并由此出发，探讨企业行为的意义。另一方面，“管理”（managerial）文献亦如雨后春笋。这类文献认为企业有其他的目标，如“满意度”、“销售最大化”，以及“管理人员效用最大化”。第二类分析框架批评第一类分析框架是不现实的，而第一类框架认为自己为企业行为提供了最优的“一阶近似”。如果企业没有使股市价值最大化，或背离价值最大化目标很远，那么就会有人试图收购它们，并令这些企业改弦易辙，从而使企业获得纯粹的资本收益。本文融合了这些截然不同的观点，提供了一个分析企业行为的统一框架。

现代公司的诸多特征通常都被忽略了，而本文将在分析中认真考虑这些特征。这些特征包括：公司有股东大会；股东大会上的投票权具有一定市场价值（有投票权的股份和无投票权的股份通常有不同的市场价格）；股东大会上经常有意见分歧；其他企业出价收购屡见不鲜；收购者往往饱受争议。现代公司是一种经济组织，总是有潜在的政治性因素（例如，投票）。因此，在我们的模型中，企业的行为由股东大会的投票决定，少数服从多数。在决定为哪个提案投票的时候，股东对于其他备选行动的后果必须有一定预期；特别地，我们只考虑小型企业，因此，个人有理由假设其他企业的行为和价值不受其影响。我们将市场均衡定义为传统的纳什非合作（Nash-non-cooperative）均衡。另一方面，企业的行动可能会对其市场价值有影响，因此，也会使得个人重新考虑其投资组合。[1]

---

* “On Value Maximization and Alternative Objectives of the Firm”, with S. J. Grossman. *The Journal of Finance*, Vol. 32, No. 2, pp. 389 - 402. 本文所受资助部分来自美国国家科学基金，资助号：SOC74 - 22382，SOC74 - 11446 - A01，以及 SOC76 - 18771，同时受添惠基金（Dean Witter Foundation）以及 IBM 公司资助。我们感谢奥利弗·哈特（Oliver Hart）给予的有益评论。

1　关于公司股东投票模型的较早的讨论，请参见 Stiglitz（1970）。

在这样的模型中，市场均衡为何样，我们将在后文中给出更详细的讨论。试想一个最简单的情形：所有人对企业应采取何种行动一致同意，在这样的情况下，股东大会在一定意义上是多余的。

文章的第一部分简要介绍了我们的结论——当股票被交易时，在很多情况下，股东们的意见显然是不一致的。第二部分建立了正式的模型。第三部分和第四部分对多期模型中关于全体一致（unanimity）的结论作了一个批判性综述。第五部分展示了如何用全体一致理论证明和分析莫迪格利安尼—米勒定理（Modigliani-Miller theorem）。第六部分和第七部分讨论了价值最大化的价值，并讨论了收购行为中的一些悖论。

一、引言：股东的全体一致

在传统的静态模型中，利润最大化的假设大受欢迎是很显然的：所有的商品和要素都有价格，企业的"价值"就是产出品的价值减去投入品的价值。所有的价格都是确知的，技术也是确知的，因此，股东们意见一致——什么样的行动可以达到价值最大化也是确定的。企业市场价值的提高，可以使个人的预算约束向外移动（平行移动，因为假设所有价格都保持不变），因而价值最大化是可取的。

如果存在期货（future）和相机要求权（contingent claim）（阿罗—德布鲁）证券的完全市场，以上结果就自然地（但并不是很有趣地）扩展到动态的、存在不确定性的情形。近期关于公司型企业行为的很多文献都可以被看作这类扩展的一种尝试——将这些关于全体一致的结果扩展到没有完全市场的情形。文献已经证实：当没有交易，并且企业的任何生产计划都是其他企业生产计划的线性组合时，全体一致是可以达到的，亦即存在所谓的"扩展生成"（spanning）。

本文的主要论点是：（i）在任何令人感兴趣的市场上都会有交易，因此，扩展生成并不意味着全体一致；特别地，我们考虑了可能产生交易的两种情形：其一，交易由于新信息的到来而产生，很显然，个人间差异化的信息可能导致交易。但是，即使新信息被所有人获得，如果个人的品味（对待风险的态度不同），均衡中的投资组合也会随新信息而变化。其二，交易由生命周期产生：在每一期，都有一些人退休，他们需要出卖股份，而又有一些人要购买股份作为储蓄为退休做准备。（ii）如果企业的财务决策也考虑在内，扩展生成本质上需要完全市场，因而也不会削弱经典的结论。（iii）除扩展生成之外，在作为扩展空间（spanned space）的一个基（ba-

sis）的组合商品市场上，假设企业为完全竞争者，那么也会产生全体一致。我们称这一假设为“竞争性”。然而，同时假设扩展生成和竞争性引出了一个很强的结论——企业欲使股份的净市场价值最大化。尽管这看似一个令人满意的结论，但却被实证结果证伪，并且未能对本绪论中所提及的诸多现象做出解释。在结论中，我们列举了一些实证的证据，这些证据来自一类特别的企业，它们毫无争议地证伪了价值最大化的假说。

二、一个简单的模型：无债务的多期模型

模型中的经济体持续三期：第0期、第1期和第2期。经济的状态 ω 由两个部分组成：在第1期所知的信号 t 和第2期所知的最终状态 s；$\omega=(t, s)$。经济中仅有一种商品，它存在于每一期中，在第0期它可被用做消费和投资，但在第1期和第2期它仅可作为消费。

在第0期，存在当期商品和企业股份的市场。同时我们假设不能签订相机商品合约（contingent commodity contract）。在第1期，也存在着当期商品和企业股份的市场。消费者在第1期消费之前获得信号 t，使得他们改变了对于最终状态 s 的信念（belief）。在第2期，根据各个消费者在第1期期末的股份，消费者获得一定的产出品。

假设存在 I 个消费者，J 个企业，T 个信号和 S 个最终状态，分别标记为 $i=1, \cdots, I, j=1, \cdots, J, t=1, \cdots, T, s=1, \cdots, S$。在下文中，有时候 I，J，T 和 S 所指的是消费者集合、企业的集合、信号的集合或者最终状态的集合。

消费者

消费者 i 的消费计划以向量 $x_i=(x_i^0, x_i^1, x_i^2)\in R_+^{1+T+TS}$ 表示，其中 $x_i^0\in R_+$ 为第0期的消费；$x_{it}^1\in R_+$ 为信号为 t 时第1期的消费；$x_i^2(\omega)=x_i^2(t, s)\in R_+$ 是 (t, s) 实现时第2期的消费。假设消费者 i 的效用函数 U_i 定义于 R_+^{1+T+TS}，并且效用函数 U_i 是严格拟凹，在其定义域内（the interior of its domain）是连续可微（differentiable）的，并且当 $x^l\to 0$，$l=0, 1$ 时，$\partial U_i/\partial x^l\to\infty$。

企业

企业 j 的生产可能性由生产集合 $Y_j\subset R_+^{S+1}$ 表示。如果 $y_j\in Y_j$，则记作 $y_j=(y_j^0, y_j^2)$，其中 $y_j^0\in R_+$ 表示第0期的投入，而 $y_j^2(s)\in R_+$ 表示第2期在状态 s 下的产出。需要注意投入品是非负的数字。我们假设 Y_j 是凸的包含原

点的闭集。进一步设定，若 $y_j \in Y_j$，并且 $y_j^0 = 0$，那么 $y_j^2 = 0$。

假设消费者 i 在第0期和第1期的初始商品禀赋分别是 $\bar{x}_i^0 \in R_+$，$\bar{x}_i^1 \in R_+^T$。同时他对企业 j 的初始持股量为 $\bar{\theta}_{ij} \geqslant 0$，其中对于每个 j，$\sum_i \bar{\theta}_{ij} = 1$。为了记号的简便，我们忽略第2期的禀赋。另外记 $\bar{X}^0 = \sum_i \bar{x}_i^0$，并且假设 $\bar{X}^0 > 0$。

固定生产计划时的均衡

假设对于每个 $j = 1, \cdots, J$，y_j 给定。第 i 个消费者通过选择（θ_{ij}，$\theta_{ij}(t)$，x_i），$j = 1, \cdots, J$，$t = 1, \cdots, T$，最优化 $U_i(x_i^0, x_i^1, x_i^2)$，受约束于个人的第2期预算约束

$$x_i^2(s,t) = \sum_j \theta_{ij}(t) y_j^2(s) \tag{1}$$

以及个人的第1期预算约束

$$x_i^1(t) + \sum_j p_j(t)\theta_{ij}(t) \leqslant \sum_j p_j(t)\theta_{ij} + \bar{x}_i^1(t) \tag{2}$$

（期末（交易后）的消费加上所持股份的价值一定不多于期初（交易前）所持股份的价值加上禀赋）；还有消费者在第0期的预算约束：

$$x_i^0 + \sum_j p_j\theta_{ij} \leqslant \sum_j \bar{\theta}_{ij}(p_j - y_j^0) + \bar{x}_i^0 \tag{3}$$

其中 $x_i^2(s, t)$ 是 x_i^2 的第（s，t）个元素；$\theta_{ij}(t)$ 为第1期信号为 t 时，对企业 j 的意愿持股数量；$x_i^1(t)$ 是 x_i^1 的第 t 个因素。$p_j(t)$ 是第1期信号为 t 时企业 j 的价格。p_j 是第0期企业 j 的价格。θ_{ij}是第0期对企业 j 的意愿持股数量。在式（3）中，我们假设购买投入品的 y_j^0 的资金全部来自于股份发售。

如果我们允许卖空（short sales）（即并不要求 θ_{ij}，$\theta_{ij}(t)$ 为非负），那么，最大化的充分必要条件就是：存在乘数（λ_i^0，λ_i^1），$\lambda_i^0 \in R_+$，$\lambda_i^1 \in R_+{}^T$，满足以下条件：

$$\nabla_0 U_i = \lambda_i^0, \quad \nabla_1 U_i = \lambda_i^1 \tag{4}$$

$$\lambda_i^0 p_j = \sum_t \lambda_{it} p_j(t), \quad j = 1, \cdots, J \tag{5}$$

$$\nabla_{2t} U_i \cdot y_j^2 = \lambda_{it} p_j(t), \quad j = 1, \cdots, J, \quad t = 1, \cdots, T \tag{6}$$

其中 $\nabla_0 U_i \equiv \partial U_i(x_i) / \partial x_i^0$，$\nabla_{2t} U_i \equiv [\partial U_i(x_i) / \partial x_i^2(t, s)]_{s=1}^S$，并且 λ_{it} 是 λ_i^1 的第 t 个元素。

从式（4）可以看出，拉格朗日乘数（Lagrange multipliers）λ_i^0 和 λ_i^1 是第0期和第1期之间消费的边际效用。式（5）可以写作 $p_j =$

$\sum_t (\nabla_{2t}U_i/\nabla U_t) \cdot y_j^2$，其含义也一目了然：第 0 期的价格等于第 2 期收益的加权平均值，其权重为第 0 期和第 1 期信号 t 下商品的边际替代率。

式（6）表明，在状态 t 下股份的价格（价值）等于再购买一股股票所带来的边际效用，并以状态 t 下收入的边际效用来标准化。

相对于生产计划（y_j），该经济的竞争交换均衡为（x_i），（θ_{ij}），（$\theta_{ij}(t)$），（p_j）和（$p_j(t)$）的集合，使得式（4）~式（6）对于每个消费者和每个商品市场在第 0 期和第 1 期都成立，并且股票市场在第 0 期和第 1 期都出清，即：

$$\sum_i x_i^0 + \sum_j y_j^0 = \overline{X}^0, \sum_i \theta_{ij} = 1, \sum_i \theta_{ij}(t) = 1 \tag{7}$$

$$\sum_i x_i^1(t) = \sum_i \overline{x}_i^1(t), \forall t \in T$$

在上述经济中，企业在第 0 期固定投入，并在第 2 期实现产出。在第 1 期，消费者获得关于第 2 期产出分布的新信息，因此，它们有机会改变在第 0 期的股份持有量。在第 2 期状态 s 下，每个消费者根据他所持有的股份比例获得产出。

在上述均衡的定义中，所有企业的生产计划都是外生给定的。在此，我们将讨论决定生产计划和交换均衡的几种方法。首先，我们将企业生产计划的选择参数化。记 $\tilde{y}_j(\rho_j) = y_j$，假设 y_j 的微小变化引起 ρ_j 的微小变化。[2]

在本文的分析中，最重要的部分就是考虑企业 j 的股东如何看待从 ρ_j 到 $\rho_j + d\rho_j$ 的微小变化。为此，我们对消费者的效用函数相对于 ρ_j 作全微分，并且定义：

根据式（1）~式（3） $U_i^*(\rho_j) \equiv \max_{(x_i, \theta_{ij}, \theta_{ij}(t))} U_i(x_i)$

通过直接计算，我们可以得到：

$$\frac{dU_i^*}{d\rho_j} = \left\{\lambda_i^0(\overline{\theta}_{ij} - \theta_{ij})\frac{dp_j}{d\rho_j} + \sum_{t=1}^{T} \lambda_{it}(\theta_{ij} - \theta_{ij}(t))\frac{dp_j(t)}{d\rho_j}\right\} + \sum_t \theta_{ij}(t)\nabla_{2t}U_i \frac{dy_j^2}{d\rho_j} - \lambda_i^0\overline{\theta}_{ij}\frac{dy_j^0}{d\rho_j} \tag{8}$$

在推导式（8）的过程中，我们用到了式（4）~式（6），以及如下假设：

$$\frac{dp_k(t)}{d\rho_j} = \frac{dy_k}{d\rho_j} = \frac{dp_k}{d\rho_j} = 0, \text{其中 } k \neq j, \forall t \in T \tag{9}$$

---

2 假设映射 $\tilde{y}_j$ 可以解出，并且交换均衡价格是 ρ_j 的可微函数。

也就是说，当第j个企业改变它的生产决策的时候，其他企业的生产决策和价格不会发生变化。（这里我们已经对企业间的竞争程度作了假定，或者隐含地对企业行为设定了纳什均衡）。如果式（8）对所有个人的正负号相同，则意味着个体对于其他生产计划的不同评价都存在全体一致。问题是，在什么时候式（8）才会对所有个人有相同的正负号。我们的目标是为了证明，在两期模型中推导出全体一致的“扩展生成”假设，在三期模型中不是充分条件。我们即将证明在“扩展生成”之外，加上一个更强的假设——“竞争性”（即将在下文中定义）——是两期模型（也就是，一个交易时期）中推导出全体一致的充分条件。

三、仅有一个交易时期下的全体一致

假设我们关闭第1期的股票市场，或者尽管市场开放，但所有交易者选择不交易。在这种情况下，企业在第0期作决策，在第2期消费者根据他们在第0期持有的企业股份比例，获得企业在第2期实现的产出。也就是说，对于所有的t，$\theta_{ik}=\theta_{ik}(t)$都成立，因此式（8）可以写成：

$$\frac{dU_i^*}{d\rho_j}=\lambda_i^0\left[(\bar{\theta}_{ij}-\theta_{ij})\frac{dp_j}{d\rho_j}-\bar{\theta}_{ij}\frac{dy_j^0}{d\rho_j}\right]+\theta_{ij}\sum_t\nabla_{2t}U_i\frac{dy_j^2}{d\rho_j}\qquad(10)$$

公式（10）右边的第一项表示财富效应（wealth effect，例如，资本利得），而第2项表示消费效应（consumption effect，例如由于第2期消费组合的改变而产生的效用变化）。

第一项表明资本利得仅在个体是证券售出者时才有意义，例如，个体的初始禀赋$\bar{\theta}_{ij}$超过了他在期末意愿持有的股票数量。因为我们特别假设了投资增加的资金来源，即来源于原始股东支付的资金，所以式（10）中有$\bar{\theta}_{ij}(dy_j^0/d\rho_j)$一项。

在不同的情形下，消费效应和资本利得效应的区别相当明显：假设只有一个汽车生产商，并且出于技术考虑，所有的汽车都是或是蓝色的或是粉红色的。同时假设较多人喜欢蓝色的车，因此，厂商可以通过全部生产蓝车来最大化自己的价值。但如果股东也购买汽车，并且知道企业的生产决策会影响他的“消费集”（consumption set），也就是，假如企业只生产蓝色汽车，则他就无法购买到粉红色汽车。那么对粉红色车有强偏好的股东就会投票要求生产粉红色汽车，即使这会导致企业市场价值的巨大损失。在股票市场上，只要没有一个相机要求权的完全市场，预算约束线的斜率总是有改变的可能。尽管并非像前例中那样的剧烈变化（粉红色和蓝色两种车的隐含的

相对价格从无穷变到零），但仍然是以一种显著的方式发生了改变。

由此，对全体一致的条件的寻求就可以分解为寻求导致出现以下三种情况的条件：（a）对消费效应的正负号存在全体一致；（b）对财富效应的正负号存在全体一致；（c）对两者的相对重要性存在全体一致。

假设存在一个证券的完全集合；则 $\sum_t \nabla_{2t} U_i = \lambda_i^0 q(s)$，其中 $q(s)$ 是状态 s 下收入的市场价格。利用式（10）和式（4）至式（6），我们得到：

$$\frac{dU_i^*}{d\rho_j} = \lambda_i^0 \left[(\bar{\theta}_{ij} - \theta_{ij}) \frac{dp_j}{d\rho_j} - \bar{\theta}_{ij} \frac{dy_j^0}{d\rho_j} + \theta_{ij} \sum_s q(s) \frac{dy_j^2(s)}{d\rho_j} \right] \quad (11)$$

需要注意此时的消费效应对所有个人都一样，另外，在这种情况下

$$\frac{dp_j}{d\rho_j} = \sum_s q(s) \frac{dy_j^2(s)}{d\rho_j} \quad (12)$$

企业市场价值的变化就是以市场价格表示的投入品和产出品的价值变化。θ_{ij} 项抵消了，我们得到 $dU_i^*/d\rho_j = \lambda_i^0 \bar{\theta}_{ij}$ [（$dp_j/d\rho_j$）－（$dy_j^0/d\rho_j$）]，即为净市场价值的变化量，这是标准的结果。然而需要注意分析过程有两个不同的阶段。第一阶段涉及对消费效应的取得的共识，第二阶段涉及比较消费效应和资本利得效应的大小。因此，如存在税收，使得消费价格和生产价格存在差距，那式（12）就无法得到；从而关于消费效应的全体一致并不意味着存在关于整体的全体一致（overall unanimity）。

显然，存在证券的一个完全集合并不是获得关于消费效应全体一致的必要条件。斯蒂格利茨（1970）注意到了获得关于消费效应全体一致的两个条件：（a）如果存在乘数不确定性（multiplicative uncertainty），那么（合适的参数化后）$dy_j^2/d\rho_j = y_j^2$，我们可以直接从式（4）和式（6）获得 $\theta_{ij} \sum_t \nabla_{2t} U_i (dy_j^2/d\rho_j) = \theta_{ij} \sum \nabla_{2t} U_i \cdot y_j^2 = \lambda_i^0 \theta_{ij} p_j$。因而对于企业 j 的所有股东，消费效应的符号独立于 i；（b）在均值方差模型中（二次的效用函数或联合正态分布的随机变量），他也证明了消费效应的符号是独立于 i 的。

艾肯—威尔逊（Ekern - Wilson，1974）和乐蓝德（Leland，1974）提供了一些一般化的条件；在这些条件下，股东对于消费效应的符号取得全体一致。艾肯—威尔逊假设企业生产计划的任何可行改变都可写为全部企业当前生产计划的线性组合；当 $\{y_j^2\}_{j=1}^J$ 给定时，他们假设存在实数 $\hat{\alpha}_{jk}$ 使得：

$$\frac{dy_j^2}{d\rho_j} = \sum_{k=1}^J \hat{\alpha}_{jk} y_k^2 \quad (13)$$

如果式（13）对 $\{y_j\}_{j=1}^J$ 都成立，我们说在 $\{y_j\}_{j=1}^J$ 上存在扩展生成。以式（5）和式（6）衡量式（10）中的消费效应，可以得到

$$\theta_{ij}\sum_t \nabla_{2t}U_i \frac{dy_j^2}{d\rho_j} = \theta_{ij}\sum_t \nabla_{2t}U_i \sum_k \hat{\alpha}_{jk}y_k^2 = \lambda_i^0\theta_{ij}\sum_k \hat{\alpha}_{jk}p_k \tag{14}$$

此式的符号独立于企业 j 的所有股东 i。不那么严格地讲，如果企业的任何产出计划都是其他企业产出计划的线性组合，那么，所有的交易者必定对产出变化的消费效应取得共识，因为消费机会并没改变。

正如我们所强调的，消费的全体一致并不意味着整体的全体一致，因为后者还必定要求关于消费效应和资本利得效应的相对重要性取得全体一致。结果表明，为了获得后者，仅仅需要以下三组条件：（a）没有资本利得的效应；（b）没有消费效应；（c）除扩展生成之外，我们还假设存在竞争性（下文将给出正式定义）。

为了考察在何种条件下可以得到（a）或（b），斯蒂格利茨（1970）指出有必要将公司的生命周期分为两个阶段；在第一个阶段，为数不多的几个投资者拥有整个企业，他们想要筹资扩大企业，每个人都希望投资于其他证券，以使投资更加多元化。在这一阶段，原始股东的原始持股量 $\bar{\theta}_{ij}$ 大大高于意愿持股量 θ_{ij}。这样，假设 $\theta_{ij}\approx 0$，我们可将式（10）写为 $dU_i^*/d\rho_j \approx \lambda_i^0\bar{\theta}_{ij}\ [(dp_j/d\rho_j) - (dy_j^0/d\rho_j)]$，即消费效应可以忽略；很显然，在此情形下，所有股东都希望企业净市场价值最大化。

在第二阶段，即企业生存的成熟稳态（steady state）阶段，股东仅在每期持有股票投资组合；例如，若市场上没有新的信息或其他冲击，稳态[3]即为 $\theta_{ij}=\bar{\theta}_{ij}$，那么，资本利得效应也消失了，我们仅仅得到消费效应（显然，如果在第 0 期有额外投资，我们必须将其成本考虑在内）。艾肯—威尔逊（1974）、乐蓝德（1974）和拉德内（Radner，1974）假设了 $\theta_{ij}=\bar{\theta}_{ij}$ 和扩展生成，以证明全体一致；并且将这一情形称为事后全体一致，以区别于事前全体一致（即这一约束不存在时的情形）。

为更精确地了解其本质，有必要考虑一个特殊的情形。假设有一个均值方差模型，以及一种无风险的资本，其第 2 期的收益定义为 R；考虑一个企业，它的回报独立于其他所有的企业，并考虑它的生产计划发生了变化，使得企业的产出在所有状态下都成比例地增加：$d\mu_j/d\rho_j=\mu_j$，$d\sigma_j/d\rho_j=\sigma_j$；这

3　然而，斯蒂格利茨很小心地指出，无交易的假设是不合理的，该假设只是为了获得对问题的分析性结论。

里 μ_j 和 σ_j 是第 j 个企业回报的均值和标准差。

假设所有交易者都有常绝对风险厌恶系数（constant absolute risk aversion），那么基本的市场估价方程（basic market valuation equation）意味着：存在一个正的常数 k，使得 $p_j=(\mu_j-k\sigma_j^2)/R$。从最优化问题的一阶条件可得，有一个正的乘数 λ_i^*，使得

$$\frac{dU_i^*}{d\rho_j}=\lambda_i^*\left[R(\bar{\theta}_{ij}-\theta_{ij})\frac{dp_j}{d\rho_j}+\theta_{ij}Rp_j\right]-\lambda_i^*R\bar{\theta}_{ij}\frac{d\bar{y}_j^0}{d\rho_j} \tag{15}$$

方括号中的第二项是消费效应，所有股东对此全体一致。然而，我们现在证明，由于资本利得这一项，股东们不一定会对生产计划取得全体一致。因为 k 是一个常数，并与股东的风险厌恶系数相关，这样就有

$$\frac{dp_j}{d\rho_j}=p_j-\frac{k\sigma_j^2}{R} \tag{16}$$

利用式（15）和式（16），可得到 $dU_i^*/d\rho_j=\lambda_i^*\{R\bar{\theta}_{ij}[(dp_j/d\rho_j)-(dy_j^0/d\rho_j)]+\theta_{ij}k\sigma_j^2\}$。因此，所有的股东都希望将企业扩张至净价值最大化的最优规模之上，因为 ρ_j 使得 $dp_j/d\rho_j=dy_j^0/d\rho_j$ 时，净价值得到最大化。

注意，我们利用了资本资产定价模型（capital asset pricing model）来计算股东如何看待企业风险回报的变化对企业价格的影响。从式（16）可看出，股东们认为讲任何状态下企业的产出翻一番不能够使企业的价值翻一番。如果在每个状态下，产出都是可分商品，那么"完全竞争"意味着每种状态下产出翻一番可以使企业的价值翻一番。如果在式（16）中，我们已经作了如此假设，那么就有 $dp_j/d\rho_j=p_j$；这样就会对净价值最大化目标取得全体一致，正如式（15）所描述的一样。

由此，我们可以得出，当交易存在时，即 $\theta_{ij}\neq\bar{\theta}_{ij}$，全体一致不一定存在。乐蓝德（1973，p. 16）和拉德内（1974）证明，如果我们假设"竞争性"，那么就会出现全体一致。竞争性意为：每个消费者都相信，任一企业的产出在任何状态下增加 $b\%$，那么，该企业的价值就会增加 $b\%$。该假设引出了隐含价格，在这些价格下，所有消费者对生产计划的改变所带来的价值取得一致意见。从式（10）到式（15），竞争性意味着对于净价值最大化目标的全体一致。

四、多期经济

在上一部分中，我们假设仅存在一期交易时间，以对关于全体一致的文献作一个评述。在这一部分中，我们重新回到第二部分的模型，其中消费者

可以在第 1 期进行交易。在第二部分的式（8）中，消费者对企业 j 的生产计划改变 $d\rho_j$ 的偏好取决于三项因子：第一项因子是第 0 期的财富效应，第二项因子是第 1 期的财富效应，第三项因子是消费效应（式（8）中的第四项是新投入品的成本所产生的财富效应）。我们可以利用扩展生成的假设（式（13）），以及一阶条件（式（5）和式（6）），将式（8）写成

$$\frac{dU_i^*}{d\rho_j} = \lambda_i^0\left[\bar{\theta}_{ij}\left(1 - \frac{dy_j^0/d\rho_j}{dp_j/d\rho_j}\right) - \theta_{ij}\right]\frac{dp_j}{d\rho_j} + \sum_{t=1}^{T}\lambda_{it}(\theta_{ij} - \theta_{ij}(t))\frac{dp_j(t)}{d\rho_j} + \sum_{t=1}^{T}\lambda_{it}\theta_{ij}(t)\sum_{k=1}^{J}\hat{\alpha}_{jk}p_k(t) \tag{17}$$

其中我们利用了隐含的竞争性假设：对于 $k \neq j$，$dp_k/d\rho_j = 0 = dp_k(t)/d\rho_j$。请注意 $\hat{\alpha}_{jk}$ 并不依赖于信号 t，因为所有企业的产出都不依赖于 t。在这个模型中，企业必定在第 0 期确定它们的生产计划，所有的企业直到第 2 期才会有产出。消费者可以在第 1 期根据新信息进行交易。从式（17）可以明显看出，如果 $\bar{\theta}_{ij} = \theta_{ij} = \theta_{ij}(t)$，那么，就会出现全体一致。然而，即使在艾肯—威尔逊—乐蓝德的假设（$\bar{\theta}_{ij} = \theta_{ij}$）下，当 $\theta_{ij} \neq \theta_{ij}(t)$ 时，一般情况下也不会有全体一致。

在一种情况下会出现全体一致，即假设竞争性。换句话说，我们可以就第 1 期的特征构建阿罗—德布鲁经济。如果把第 1 期信号 t 下的企业价值写为一系列特征的混合（即一种合适的证券组合），且每一特征都有一个竞争市场，那么，每个信号状态下企业价值都可以像阿罗—德布鲁经济中那样计算得出。格罗斯曼—斯蒂格利茨（1976a）证明，在合适的竞争性假设下，总是可以做出上述推导，并且对于价值最大化的目标能有全体一致。这样，我们证明了当经济是阿罗—德布鲁经济的一个拓展时，就算有交易行为，也总能得到全体一致。

正如第三部分所述，无交易假设对于任一全体一致理论都很重要；全体一致理论并非阿罗—德布鲁证券市场的一个简单拓展。在本部分中，我们证明新的信息可以产生交易，个人的生命周期同样也会产生交易（我们之前所讨论的企业多个阶段是股票市场上交易持续产生的第三个理由：企业也有生命周期）。当某些个人年轻时，他们购买股票作为储蓄的手段；当他们变老时，便出售他们的股票，一般是出售给下一代人。

假设个人的寿命为两期。在第 1 期，他们工作并且购买股票；在第 2 期，他们获得分红（决定于企业上期的行动），并做出决策——这决定了下期的红利，出售股票，并最终获得收益。在这样的情况下，很显然，在决策

制定之际，对于决策者而言 $\theta_{ij}=0$。这样，所有的企业都会选择让股市价值最大化。但是请注意，如果我们将模型改为三期模型，那么在第 1 期和第 3 期都会有持股者。对于前者而言，$\theta_{ij}>0$，而对于后者，如前面一样，$\theta_{ij}=0$。这样的话，我们不能保证有全体一致，除非扩展生成和竞争性假设得到满足。如果不存在交易，仅有扩展生成假设也已足够。

五、扩展生成和债务—权益比例的选择

在此，我们回到第三部分的一期交易模型。假设生产计划 $y_j=(y_j^0, y_j^2)$ 已经选定，企业现在的问题是如何获得购买 y_j^0 的资金。记 D_j 为第 0 期企业 j 发行的债务的价值。记 $1-a$ 是原始股东为筹资而发售的股份（即新的股票）的比例。那么 $y_j^0=D_j+(1-a)p_j$。记 B_j 是企业承诺在第 2 期付给债券持有者的资金总额（本金加利息）。记 b_{ij} 为消费者 i 所持有的企业 j 的债券比例。这样，式（1）可以写成 $x_i^2(s)=\sum_j\theta_{ij}\max(0,y_j^2(s)-B_j)+\sum_j b_{ij}\min(B_j,y_j^2(s))$。消费者在第 2 期的消费来自于得到偿付的债券收入和对产出品的剩余索取权。类似地，式（3）可重写为 $x_i^0+\sum_j p_j\theta_{ij}+\sum_j D_j b_{ij}\leqslant\sum_j\bar{\theta}_{ij}(p_j-y_j^0+D_j)+\bar{x}_i^0$。这样，每个原始股东拥有企业 $a\,\bar{\theta}_{ij}$ 比例的股份，其价值为 $a\,\bar{\theta}_{ij}p_j=\bar{\theta}_{ij}\ (p_j-y_j^0+D_j)$。

我们假设在第 1 期没有交易；这样，消费者通过选择 x_i^0，θ_{ij} 和 b_{ij}，在上述约束下最大化 $U_i(x_i^0, x_i^1, x_i^2)$。记最大化的效用为 $U_i^*(B_j)$；同时 $B(B_j)=\{s\mid y_j^2(s)<B_j\}$，$NB(B_j)=\{s\mid y_j^2(s)>B_j\}$。我们感兴趣的是，要保持生产计划 y 不变，债务—权益比例应该如何变化。我们可以在 B_j 点（且 $y_j^2(s)\neq B_j$ 对所有 s 都成立）计算 $dU_i^*(B_j)/dB_j$。这由下式给出：

$$\frac{dU_i^*(B_j)}{dB_j}=\lambda_i^0(\bar{\theta}_{ij}-\theta_{ij})\frac{dV_j}{dB_j}+(\theta_{ij}-b_{ij})\left[\frac{dD_j}{dB_j}-\sum_{s\in NB}\nabla_s U_i\right] \tag{18}$$

其中 $V_j\equiv D_j+p_j$，并且 $\nabla_s U_i\equiv\sum_{t=1}^{T}\partial U_i(x_i)/\partial x_i^2(t,s)$。

莫迪格利安尼—米勒定理给出了在哪些条件下债务—权益比率的变化不会改变企业的价值。然而，更严密的理论必然要求所有股东对债务—权益比例都毫不在意。从式（18）可以看出，$dV_j/dB_j=0$ 并不等同于 $dU_i^*/dB_j=0$。只有当 $\theta_{ij}=b_{ij}$时（这一条件很重要，并且是广为人知的），两者才是等同的。$\theta_{ij}=b_{ij}$出现的条件是存在投资组合可分（portfolio separation），使得所有

消费者希望持有相同的共同基金，以此投资于所有的风险资产（例如，所有消费者一致认为，所有证券的回报服从多元正态分布），也就是说，以相同比例持有所有的风险资产。

我们感觉莫迪格利安尼—米勒定理的要点在于债务—权益比率是不确定的，因为所有的股东对于 B_j 的选择全都毫不在意。要使这个条件成立，必然要求 $dU_i^*/dB_j=0$ 在其有意义的定义域成立。从式（8）可以看出，一般来说这不成立。使其成立的一种情形是：不存在交易，即 $\bar{\theta}_{ij}=\theta_{ij}$；以及存在强的投资组合可分，即 $\theta_{ij}=b_{ij}$。但是，如同事后全体一致不能推广至多期模型一样，这个结论也不能推广至多期模型。

读者可能会认为，扩展生成假设可以在上述情形中推导出全体一致。这是正确的，但它也意味着在较弱的正则条件（regularity conditions）下存在着完全市场，请参见格罗斯曼—斯蒂格利茨（1976a）所作的证明。

六、价值最大化的价值

在上文中，我们已尝试证明了大部分扩展生成文献中的全体一致理论都需要假设：在企业产出的实现和购置投入品之间的那段时间没有股票交易。也就是说，简单的扩展生成理论关注不在大的交易所交易股票的企业。在这些大的交易所交易的股票，每天都有很大的交易量，消费者能获得关于企业产出概率分布（即红利）的新信息。我们已经证明，在这样的经济中，消费者通常不会对企业应采用的生产计划取得一致。这些不一致的意见，是因为不同的股东在新信息来临时对资本收益和损失的估计不同。

如果在扩展生成假设之外加上竞争性假设，我们就可以在多期模型中证明全体一致理论。尽管竞争性假设看似一个微不足道的额外假设，但它却可推导出扩展生成假设所不能推导出的很强的结论。也就是说，如果存在扩展生成和竞争性，那么，股东们会一致希望企业最大化其净市场价值。

正是价值最大化的含义，使得我们认为这一理论存在严重的缺陷。市场上显然存在着一类企业，它们的净价值未得到最大化。封闭式共同基金是在股票市场上购买其他企业股票的一种企业。此类基金的唯一生产决策是购买或者出售某一企业的股票，从而获得现金或者购买其他企业的股票。而这类基金本身也是一类公司，它们的股票也在股市上交易。通常所有的封闭基金出售时都有很大的折扣（Sharpe-Sossin）。即共同基金投资组合的市场价值大大高于基金本身股份的市场价值。这表明基金经理可以做出一种决策以提升企业价值，即经理可以出售股票投资组合换得现金，并将现金分配给共同

基金的股东。因为事实上没有人这样做，我们认为共同基金的股东并不希望基金让市场价值最大化。

关于全体一致的大部分讨论涉及生产决策的选择。我们认为莫迪格利安尼—米勒定理成立条件是：不存在破产，这样所有股东对债务—权益比率的选择全都毫不在意。我们证明了，如果企业有可能破产，股东们不一定对债务—权益比率全都毫不在意。因为不同企业相同期限的债券有不同的收益，所以，在现实世界中企业必定有破产的可能。读者可能会认为如果存在债券的扩展生成，莫迪格利安尼—米勒定理还是成立的。然而，债券的扩展生成意味着存在完全市场；这当然也就意味着债务—权益比率是不确定的。但这也意味着企业让其价值最大化，并且所有风险都是可以买保险的——这两个都已经被实证证伪。

七、公司型经济中的均衡

我们的分析表明，通常不存在全体一致，那么，这会带来什么后果呢？如果扩展生成假设得到了满足，而竞争性假设未得到满足，因此，全体一致存在，但非价值最大化策略的全体一致，那又会怎样？会不会有人收购该企业，改变政策，实行市场价值最大化政策，以获得纯粹的资本利得？回答这些问题并且构建一个完整的市场均衡理论，需要对股东拥有的信息、收购行为以及个人对收购后所采取行动的预期做出特别的假设。我们认为收购的均衡理论构建过程中存在严重的问题（参见格罗斯曼—斯蒂格利茨（1976）关于没有扩展生成的企业理论）。

假设对于某一政策（但非价值最大化）存在全体一致，再假设收购了企业的个人宣告：该企业会实行价值最大化策略；所有人都知道，在收购后，收购者将会出售他的股份以获得资本利得，并且在其后的股东大会上，所有股东都会投票赞成原先的效用最大化策略。这样，尽管有那样的宣告，所有人都预期企业不会实行价值最大化策略，因此，收购者也不能得到任何资本利得。在上述例子中，如果收购者可以让人相信企业会在一期中实行价值最大化策略（比如，因为固定投资），那么，他就可以在当前市价水平上获得多数股份。但除非竞争性假设得到满足，对企业股份的需求可能是一条向下倾斜的曲线。市场价值代表了边际购买者的估价；为购买大部分股份，即使采用歧视性垄断者的做法，其成本也会大大高于企业的现值。

即使所有股东都希望价值最大化，如果当前管理层不愿实行这样的政策，要想实行一个成功的收购还是会有很多问题的：（a）如果与收购者在

收购前宣布了收购计划，那么，其他人出售股票就不合算了；因为如果其他人预期收购会成功，那么，通过延期出售股票，他们可获得资本利得。或许收购者可以以最大化的市场价值收购股份，但这样的话，收购行为便无利可图了。（b）如果收购没有提前宣布，而是逐步实施，只有当收购的股份达到10%的时候，收购行为才间接宣布（正如现行法律规定），这样也会产生问题。如果成功的收购与市场价值的增加相结合，那么在收购宣布后，市场认识到这点，股价就会立刻上涨。但如果这样的话，任何人都会倾向于购买10%的股份，以便获得资本利得。到那时，当然宣告（disclosure）也不再是收购的信号了。收购有许多格罗斯曼—斯蒂格利茨（1976a）所讨论的套利的信息特征。

我们现在已经可以给出一个均衡的定义。考虑所有具有如下特征的政策：当这一政策宣布时，大部分购买股票的持股者[4]会投票赞成此项政策，而投票反对其他政策。这样的每一组政策都有市场价值。均衡要求这项政策得到大部分股东的支持，并且没有被收购之忧。具有最高市场价值的政策就是均衡。显然，其他各项政策都不能成为均衡；为忽略购买股票的两难困境，宣布了均衡政策的收购者可以维持这项政策，因而也可以获得资本利得。如果所有个人假定企业最终实施的政策——因而也使相关的政策——是可以在上述意义下维持的政策，那么，其他任一政策所带来的市场价格都会比这一市场价格低，因此在此政策下，不会有人试图收购。

八、结论

本文的论点总结如下：（a）扩展生成并不意味着在股票被交易的公司中的股东们会实现全体一致。（b）如果因为竞争性假设，股票被交易的企业也有全体一致，那么，这样的企业必定会最大化其价值。（c）但封闭式基金并不让其价值最大化，而它们正是我们预料最应实施价值最大化的企业。基于此，我们认为扩展生成假设不能令人满意。（d）我们进一步论证了如下论点：若要以对市值被低估企业的收购行为证明价值最大化，其中也有许多根本性的困难。

4　显然，正如在任一投票模型中一样，股市投票模型也不一定存在均衡；改变计划或者议而不决的成本也都很高；因此，我们做了如下的修正：多数投票均衡是一个提案，其他任何未输于某些提案的提案也不能在投票中赢过它。

参考文献

P. A. Diamond. [1967], "The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty," *American Economic Review*, 57, 759–776.

S. Ekern and R. Wilson. [1974], "On the Theory of the Firm in an Economy with Incomplete Markets," *The Bell Journal of Economics and Management Science*, Vol. 5, No. 1, 171–180.

S. Grossman and J. Stiglitz. [1976a], "On Stockholder Unanimity in Making Production and Financial Decisions," Technical Report No. 250, Economics Series, Institute for Mathematical Studies in the Social Sciences (IMSSS), Stanford University.

——— and ———. [1976b], "Information and Competitive Price Systems," *American Economic Review*, Vol. 66, No. 2, pp. 246–253.

——— and O. Hart [1976], "A Theory of Competitive Equilibrium in Stock Market Economies," Technical Report No. 230, Economics Series, IMSSS, Stanford University.

H. Leland. [1973], "Capital Asset Markets, Production and Optimality: A Synthesis," Technical Report No. 115, Economics Series, IMSSS, Stanford University.

———. [1974], "Production Theory and the Stock Market," *The Bell Journal of Economics and Management Science*, Vol. 5, No. 1, 125–144.

R. Radner. [1974], "A Note on Unanimity of Stockholders' Preferences Among Alternative Production Plans: A Reformulation of the Ekern-Wilson Model," *The Bell Journal of Economics and Management Science*, Vol. 5, No. 1, 181–186.

W. F. Sharpe and H. B. Sossin. [1975], "Closed-End Investment Companies in the United States: Risk and Return," in *European Finance Association, 1974 Proceedings* (B. Jacquillant, editor), North Holland, 37–63.

J. E. Stiglitz. [1970], "On the Optimality of the Stock Market Allocation of Investment," paper presented at the Far Eastern Meetings of the Econometric Society, Tokyo, June 27–29, 1970.

———. [1972], "Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Take-Overs," *The Bell Journal of Economics and Management Science*, Vol. 3, No. 2, Autumn 1972, 458–482.

关于公司财务政策的无关性*

本文将默顿·米勒和弗兰科·莫迪格利安尼以及作者（1969）关于公司财务政策无关性的命题拓展为一个多期模型。这样，我们就能够考虑更多种类的财务政策：不但包括公司选择的债务—权益比率，还包括公司选择的股利—留存收益比和债务的到期结构，甚至包括公司持有其他公司资产（证券）的决定。我要证明的是，在一个一般均衡的环境下，在相当弱的条件下，以上所有这些政策都不会对公司的价值造成影响。关于这些假设是否"符合实际"是一个有争议的问题，而关于这个问题，在本文后面部分我将阐述我的看法。但是通过对假设的阐明，至少我希望我们的讨论是在有关的问题上展开的。

对学习公司财务的学生而言，公司财务政策对公司价值影响的问题无疑是一个中心问题；如果得到"无关性"定理的条件被视为是符合实际的，那么，这将会夺去他们绝大部分交易中的股票。但是，公司财务政策的无关性问题远比这重要得多。

我们可以将公司的决策分成四类：

（1）公司如何对投资进行融资？

（2）公司如何分配收入？

（3）公司应该怎样进行投资？

（4）公司应该投资什么样的项目（或者公司应该采用什么样的生产技术）？

头两类决策是公司的财务决策，后两类是公司的投资决策。公司财务理

---

* "On the Irrelevance of Corporate Financial Policy", *The American Economic Review*, Vol. 64, No. 6 (Dec., 1974), 851 - 866. 这是作者1970年6月25 - 26日在日本箱根（Hakone）举行的一次会议上的讲演稿的修订版本。作者要非常感谢会议的组织者（Hirofumi Uzawa）和会议的其他参与者的有益建议和讨论。本文的研究受到了古根汉基金、美国国家科学基金会、福特基金会的大力支持。就像所有其他公司财务领域的研究者一样，作者本文的研究在很大程度上受益于默顿·米勒（Merton Miller）和弗兰科·莫迪格利安尼（Franco Modigliani）的思想；另外，关于本文所讨论的问题，作者从多年来与他们的有益的讨论中受益匪浅。同时作者还要感谢与罗伯特·默顿（Robert Merton）和皮特·卡贝尔（Peter Kerbel）的有益的讨论。

论着重研究公司的财务决策。公司的两类财务决策是紧密相连（见后面），而两类投资决策也同样是紧密相连。我们并不清楚公司的财务决策与投资决策之间有什么关系。关于这个问题的回答，我们需要对公司与该经济中的家庭户之间的关系进行分析，而拓展我们对这种关系的理解是本文的一个主要目的。如果关于公司的财务政策与公司的市场价值无关的假说是正确的话，那么，这也意味着，如果公司要最大化市值，那么，唯一与此有关系的决策就是投资决策，而财务决策对此并无影响。特别是，这也意味着基于"现金流分析"的真实部门分析，以及诸如尼古拉斯·卡多尔（Nicholas Kaldor）的："因为从家庭户部门流向公司部门的现金流是很小的，所以家庭户关于储蓄的决定对经济均衡的决定的影响是相对较小的"之类的结论，对于我们理解真实世界是不会有多少帮助的。它最多只能告诉我们一些貌似存在的关系。[1] 更进一步地，如果债务的到期结构是与公司价值无关的话，那么这就使得，我们对企图将到期结构与利率的期限结构联系起来进行局部均衡分析的做法（例如，见我的1970年的论文）的正确性产生了怀疑。

在理论文献中，两个紧密联系但并不相同的命题被混淆了。这两个命题都认为公司的财务政策对公司的价值没有影响。但是其中一个命题认为，人们对财务政策是无差异的，特别是对于债务—权益比率，所以对整个经济体而言，并没有最终的债务—权益比率。这就是说，任何公司的财务政策的变化都能够完全被股民的行动所抵消（而且事实上将在新的一般均衡情况中被抵消）。

第二个命题认为，也许人们对于可选择的财务政策并不是觉得无差异的，因而，整个经济体就会存在一个确定的负债—权益比率，但是对每一个特定公司财务决策是没有影响的。换句话说，第一个命题认为，对整个经济而言，财务政策是无关的，从而对某个特定的公司来说财务政策也是无关的；而第二个命题只认为，对特定公司的财务政策是无关的。显然前一个命题要比后一个命题更站得住脚。[2] 这里我们对两个命题

---

1 例如，留存收益与投资之间的关系并不能为投资水平是如何决定的提供一个解释。例如，见约翰·梅耶（John Meyer）和埃德温·库恩（Edwin Kuh）的著作。

2 莫迪格利安尼—米勒定理正是属于后一种类型。他们证明如果存在两家或者多家公司属于同一风险种类（在不同的自然状态下具有相同的收入形式），那么任何一家特定公司的负债—权益比率都是无法决定的。而作者（1969）论文的结论就是属于前者类型的。

都很关注。

本文的顺序是这样安排的。在第一部分，建立基本模型。在第二部分，我证明本文的基本定理：从任何个体的角度来看，公司财务政策都是无关的。第三部分对于定理的假设及其局限性做出简要的评价。第四部分将证明，在比前面定理的证明所用到的假设弱得多的假设条件下，即使财务政策对于个人是值得关注的，对于任意的特定的公司它也不一定是重要的。

一、基本模型

1. 公司

一个公司的各种财务决策是紧密联系在一起的。在一个多期模型中，一个很有意思的问题就是探讨它们之间的关系。增加股利派发资金量的决定意味着，如果公司保持预定的投资计划不变，那么公司就需要筹得额外收入，以支付预定的投资数量。如果公司决定通过发行债券来筹集这部分额外的资金，那么，剩下的用来作为下一期利润分配的资金减少，从而意味着要么下一期的留存收益减少，要么下一期的股利减少。如果公司通过发行新股来募集这部分资金，这就意味着下一期派发给各位股东的收益就会减少（如果下一期的留存收益不变）。由此可见，这些决定之间的相互联系是复杂的，而且今天的决定对于公司未来的影响可能会持续很多期。[3]

为了方便，我们将建立一个“一种商品”的模型；[4] 在每一期有一个唯一的投入和一个唯一的产出（美元或者日元）。我们将要考察在给定公司的一个“真实计划”情况下，不同的财务计划对公司价值的影响。一个真实投资计划的特征取决于每一期每一种可能的自然状态下（在此之前所有发生的事件的集合）投资水平和技术选择。所以，给定一个真实的计划，公司的每一期利润当然取决于自然的状态。令

3　这些关系的重要性经常被即使是公司财务理论的精英，例如，对威廉·巴莫尔（William Baumol）和伯顿·玛吉勒（Burton Malkiel），莫迪格利安尼和米勒理论的研究者们所忽视。在讨论税收对于公司的最优财务政策的影响时，他们认为，增加的债务降低了公司的税收义务，从而提高了公司的价值。对于一家采用固定投资策略的成长中的公司，提高的负债—权益比率意味着公司的留存收益将会更少，所以资本收益会更少。

4　对多商品模型的分析也是一样的，除了公司有了新的可以利用的财务决策：公司可以用钱结算它发行的债券，也可以用其他某种商品，也可以用某种复合商品来结算（例如，生活指数成本）。事实上某种看起来比较“真实”的决策正是财务决策：当相对价格存在不确定性时，公司必须决定是否购买投入品的期货（还是保持库存）和是否出售产出品的期货（还是保持库存）。简而言之，（在假设条件下）所有这些“对冲”决策都不会对公司的价值产生影响。

$I_i(t,\ \theta(t),\ k)$ =给定真实计划 k，如果在时期 t 自然状态为 θ（t），第 i 家公司在 t 期的投资水平。

$X_i(t,\ \theta(t),\ k)$ =给定真实计划 k，如果在时期 t 自然状态为 θ（t），第 i 家公司在 t 期的产出或者总利润。[5,6]

公司可选择的投资和融资方式有很多种：

（a）公司可以通过留存收益或者发行新的证券来投资筹集资金。

（b）如果通过发行新的证券，那么就有很多种不同的可供公司选择的金融工具：普通股、债券、优先股、可转换债券等。每种金融工具对公司总利润的分配和这些金融工具的所有者在公司决策中所扮演的角色等方面都有不同的影响。例如，债券在每种自然状态下都产生一笔固定的数额的收益，除非公司破产，那样的话，公司的财产将在债券所有人之间分配。但是除非当公司存在很大的可能性不能履行它的偿债义务时，债券所有人在公司的经营管理方面没有投票权。

一方面，普通股的收益是可变的，除非公司破产，如果那样的话普通股的收益就是零。一个股票持有者有权得到公司股利中与其所持股份数相对应的份额，当然，股利不但取决于公司的真实投资计划，而且还取决于公司所采取的财务政策。要了解收入流，股东不但要了解公司的真实决策，还要了解公司的金融决策。另一方面，如果我们关于公司金融决策无关性的结论是正确的，那么，虽然公司的分红政策的变化会影响到公司每一单独股份所获的收益流，但是人们对这些变化是无差异的。股票持有者（类似债券所有者的做法）可以随时出售他所持有的股票而获得收益。最后，股票的所有也意味着股票持有者在股东大会上拥有相应比例的投票权（虽然很多公司经常发行没有投票权的股票）。

在接下来的分析中，为了简单起见，我们将假设只存在两种金融工具，普通股与债券。

（c）如果公司决定发行债券，那么它必须决定债券什么时候到期——一年、两年等——以及票面利率是多少。为了简单起见，我们假

5　为了不失一般性，我们可以将 I_i 和 X_i 理解为表示投入和产出的向量。从而在证明过程中，我们需要用 $v \cdot I_i$ 去替代 I_i，用 $v \cdot X_i$ 去替代 X_i，这里 v 为价格向量（在 t 期，θ（t）状态下）。

6　因为在我们的简化模型中没有其他种类的投入品，公司的产出和它的总利润（在未支付债务利息等之前）就是一样的。当存在其他种类的投入时，需要做的改动是很明显的。

设债券没有票面利率。所以一份 t 期的债券就意味着在 t 期过后支付一美元的承诺。到该债券公开发行时，当然会折价发行。[7] 令

$p(t, \tau, \theta(t))$ ＝在 t 期，在 $\theta(t)$ 状态下，承诺在 τ 支付一美元的债券的价格。

如果存在不确定性，人们将不能确定这样一份债券在将来时期的价格，除非公司不破产的话，那么

对所有的 θ 和 t 有：$p(t, t, \theta) = 1$

在下面的讨论中，所有的变量都取决于自然状态，但是为了简单起见，如果不引起混淆，我将省略 $\theta(t)$。类似地，所有的真实变量都将取决于"计划 k"，但是 k 通常也被省略。

不同种会计决策之间的关系可以通过如下两个会计恒等式来表达：总投资必须等于在外流通的债券价值的变化，加上在外流通的股票价值的变化，再加上留存收益。即

$$I_i(t) = \sum_{\tau=t+1}^{\infty} [p(t,\tau)(B_i(t,\tau) - B_i(t-1,\tau)) + q_i(t)(S_i(t) - S_i(t-1))] + RE_i(t) \quad (1)$$

这里

$B_i(t, \tau)$ ＝在 t 期末在外流通的，τ 期到期的债券总价值；

$q_i(t)$ ＝在 t 期第 i 家公司的股票价格；

$S_i(t)$ ＝t 期末在外流通的总股份数；

$RE_i(t)$ ＝留存收益。

当然，这里并不存在股份数的自然单位，所以，我们简单地定义

$$E_i^+(t) = q_i(t)S_i(t)$$

为该期末在外流通股份的总价值，而定义

$$E_i^-(t) = q_i(t)S_i(t-1)$$

为该期初在外流通股的总价值，也就是说，前一期末在外流通的股份数在第 t 期的价格水平下的总价值。

所以 $E_i^+(t) - E_i^-(t)$ 就是：在第 t 期发行新股所导致的，在外流通的股份数的变化的价值；这不能和 $E_i^-(t+1) - E_i^+(t)$ 混为一谈，这是指，在 t 期末流通的总股份数在 t 期到 $t+1$ 期之间的价值的变化。它也可以理解为已

7　这里当然仍然存在其他的金融决策，包括对其他公司资产的所有者权益，用什么本位商品为债券计价等。所有这些都可以被纳入到本模型中，只是会增加表达上的复杂性。

有股份的资本所得（或者损失）。

第二个会计恒等式说的是，在 t 期状态 θ 下的总收入必须等于所分配的收入（包括分配给债券所有人和股东的）加上留存收益。[8]

$$X_i(t) = B_i(t-1,t) + D_i(t) + RE_i(t) \tag{2}$$

这里，$D_i(t)$ 为 t 期初记录在案的发放给股东的股利；即每股获得股利为 $D(t)/S(t-1)$，或者 $t-1$ 期每一美元的投资所获得的股利为 $D(t)\ /E_i^+(t-1)$。

图1所示的是一段时间内该经济随资金流情况（由于该资金流是沿着时间先后顺序发生的，所以该图并没有被画成圆形）。应当注意到，我们将留存收益的线画成一条通过家庭户部门的虚线：收益并不直接从家庭户手中经过的事实，并不一定意味着家庭户在他们的收入中（从某种意义上说）不包含这种留存收益。正如下面分析所说明的那样，这种留存收益将直接反映在流通股的价格上。

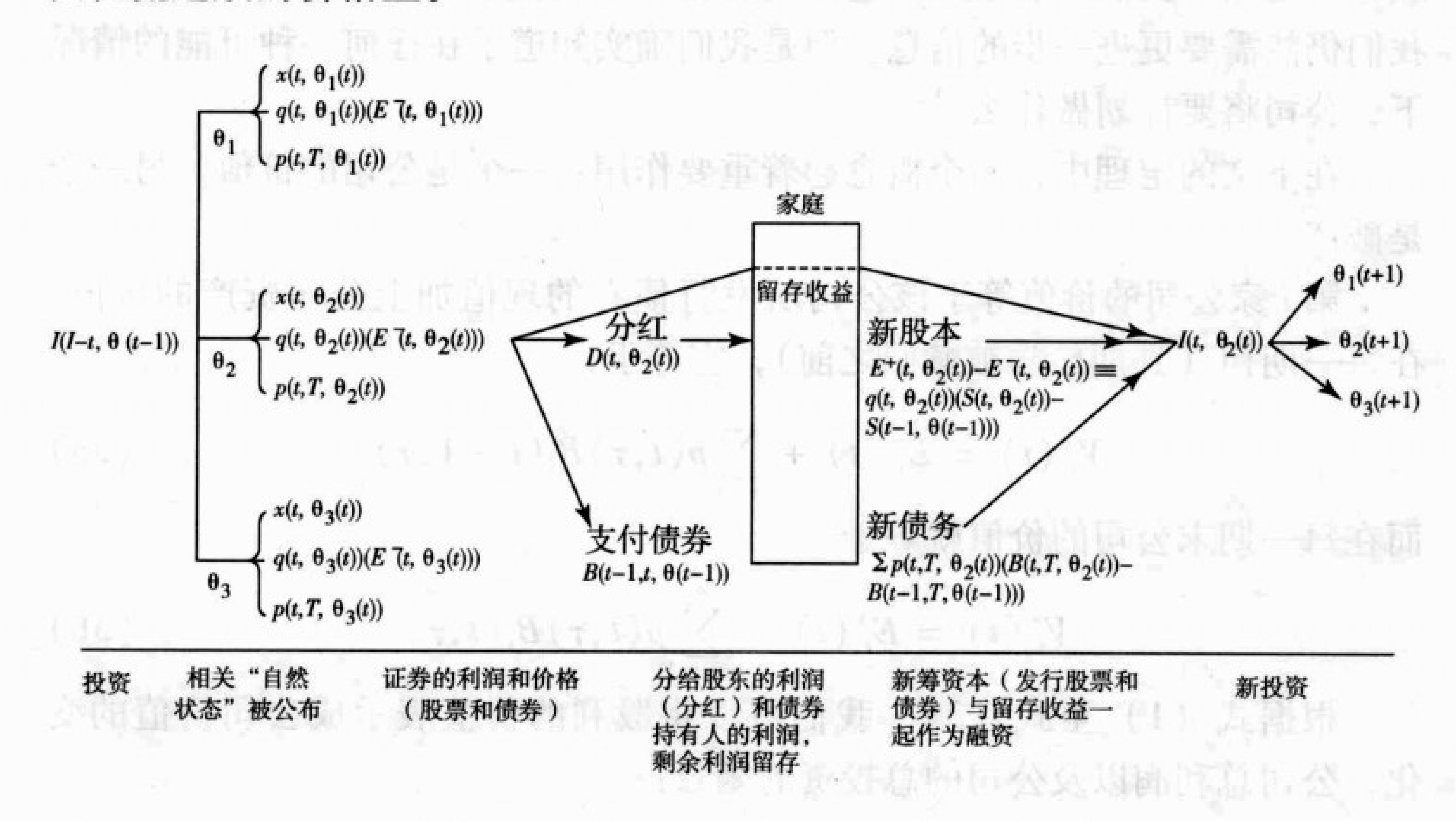


图1

这张图表也可以用来说明我们分析中隐含的时间顺序：让我们从图表中代表公司刚刚做出“新的”投资决定，正在募集所需要的资金那一点开始。下一期的产出（利润）（这不但取决于刚结束那一期的投资水平，也

8　回忆一下，我们为了符号上的简化，假设债券不存在票面利息，所以，债券持有人只通过到期债券的形式从公司获得收益。

取决于之前所有期的投资水平，同时也受到之前所有期所实现的自然状态路径的影响）是未知的；我们等待着第 t 期的“自然状态”的实现，例如，降雨、温度等。这时，自然状态被公布，即 $\theta(t)$ 被给定。从而，第 t 期及其以后时期可能结果的集合就立即被相当程度地缩小了。例如，在图 1 中，除了一条经过 t 期的自然状态路径，其他所有的路径，我们就可以完全忽略。

给定状态 $\theta(t)$ 的明确化中所包含的信息，股票的价值和债券的价格就被确定了。特别是，公司资产现在的价值为 $E_i^-(t,\ \theta_2(t))$。更进一步地，在这时候，对于我们记为 k 的特定计划，我们明确地知道公司在这一期计划做什么：我们知道它的投资水平 $I(t,\ \theta_2,\ k)$，它计划发放多少股利，留多少收益；它计划发行多少债券，债券的到期结果是什么；它计划发行多少新股等。当然，我们仍然不知道它的投资在将来会是什么状态，对于这一点，我们仍然需要更进一步的信息。但是我们确实知道了在任何一种可能的情况下，公司将要计划做什么。[9]

在下文的定理中，两个概念起着重要作用：一个是公司的价值，另一个是破产。

第 i 家公司的价值等于该公司所发行债券的现值加上公司资产的价值：在每一期初（到期债券被赎回之前），它等于：

$$V_i^-(t) = E_i^-(t) + \sum_{\tau=t}^{\infty} p(t,\tau)B_i(t-1,\tau) \tag{3a}$$

而在每一期末公司的价值就等于

$$V_i^+(t) = E_i^+(t) + \sum_{\tau=t+1}^{\infty} p(t,\tau)B_i(t,\tau) \tag{3b}$$

根据式（1）至式（3），我们可以将股利的价值表示成公司价值的变化，公司总利润以及公司的总投资的函数：

$$D_i(t) = X_i(t) - I_i(t) + V_i^+(t) - V_i^-(t) \tag{4}$$

破产可能更难定义一点。当然，基本的概念就是公司不能履行它的偿债义务。在我的 1969 年的论文中所讨论的两期模型中，只要当公司的总利润

9 假设（在每一期）募集新的资本在利润分成和债券支付之后仅仅是为了说明的方便。实际上，这两个过程可能会同时进行。重要的是，我们假设在每一期的期初和期末证券的价格水平相同。因为一旦自然状态确定了，那么这一期将要发生什么也就知道了，所以这个假设并不是一个毫无道理的。如果我们将收益的分成，新的资本金的募集理解为同时进行的话，那么这显然是个合适的假设。

小于公司债券持有人的名义应得收益

$$X_i < (1 + \hat{r}_i)B_i$$

我们就称该公司破产。这里 $\hat{r}_i$ 为公司债券的名义利率。如果 r^* 为百分之百安全债券的名义利率，而且 $\min X_i(\theta)$ 为在任意自然状态下公司的最低利润水平，公司破产概率为零，假定：

$$B_i \leqslant \frac{\min X_i(\theta)}{1 + r^*}$$

而如果

$$B_i > \frac{\min X_i(\theta)}{1 + r^*}$$

那么，概率就大于零。如图 2 所示，这里类似的结论为

$$X_i(I_i, \theta(t)) < B_i(t - 1, t, \theta(t - 1)) \tag{5}$$

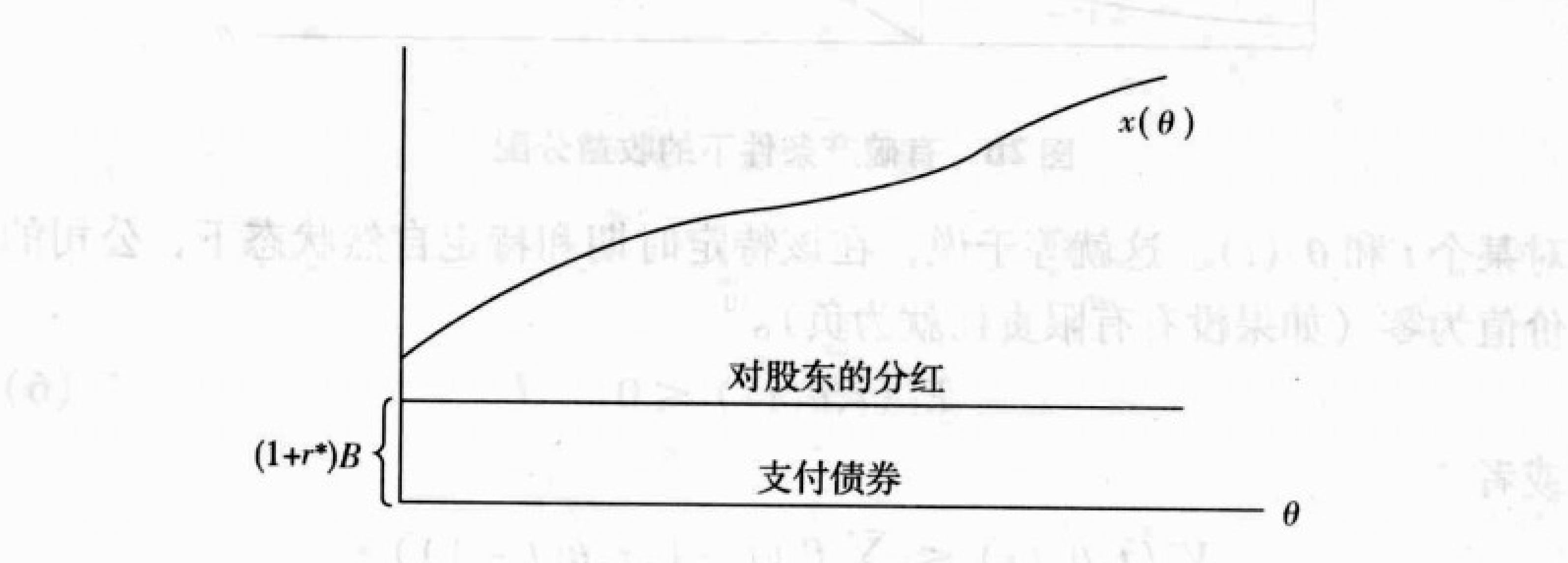


图 2a　无破产条件下收益的分配

但是这样定义还不够。因为如果公司在某一时期的特定状态下收益很低的话，公司总是可以通过借更多的钱，或者发行新股来募集履行还债义务所需的资金。事实上，假定公司的盈利前景足够的好，公司通常会那样做。因此，式（4）中所表示的关系至多是关于公司短期流动性的一种表述，而并不说明给定公司的短期偿债能力。而在一个存在最终时期的多期模型的最后一期，公司的破产条件由式（5）给出，因为此时（根据假设）公司不再可以借钱或者发行新股。但是我们的分析并没有理由限制在有限期的情况。

显然，我们所谓的破产就意味着在某一期、某一特定的自然状态下，公司的到期债券的价值要小于其面值，即

$$p(t, t, \theta(t)) < 1$$

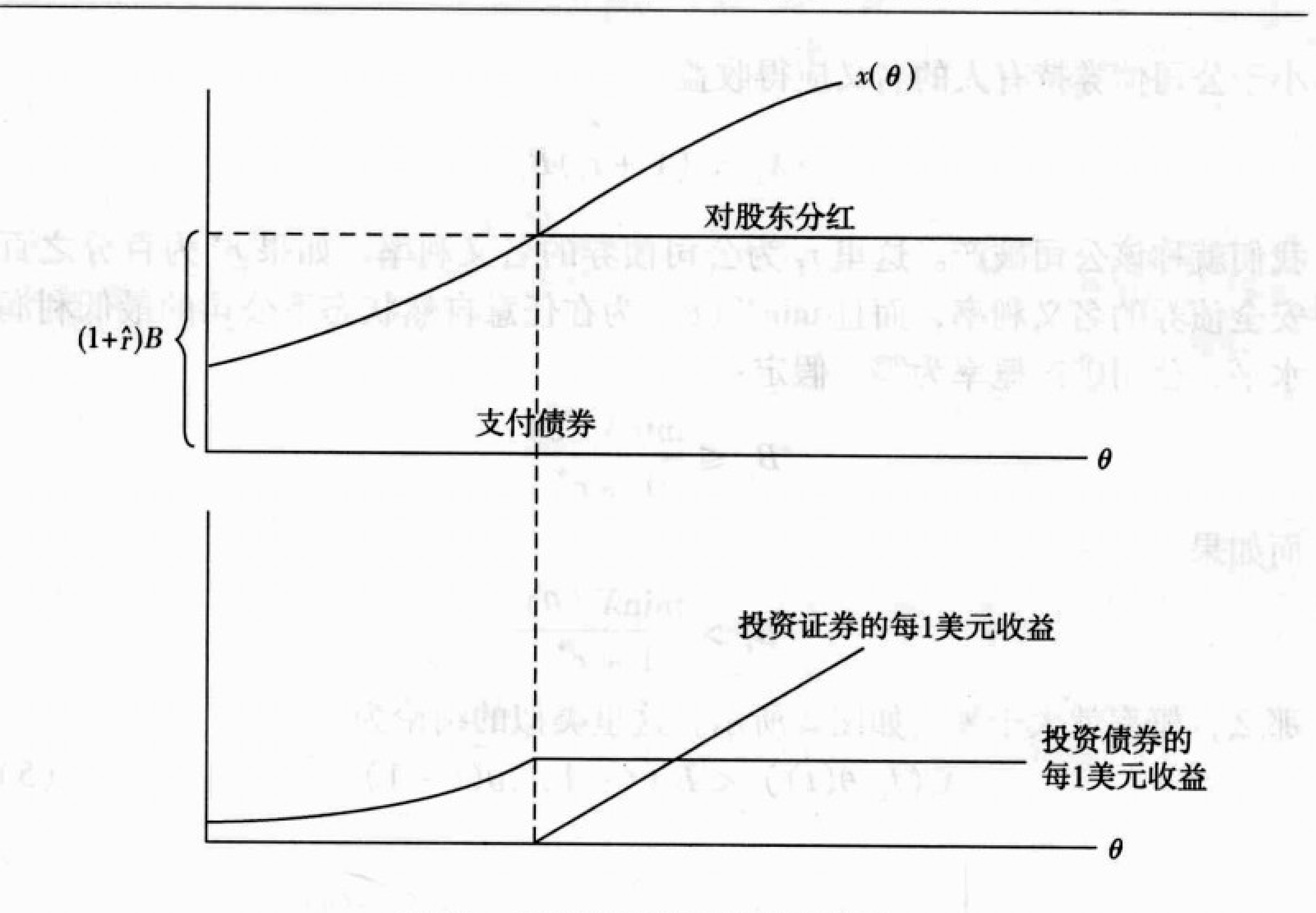


图 2b　有破产条件下的收益分配

对某个 t 和 θ（t）。这就等于说，在该特定时期和特定自然状态下，公司的价值为零（如果没有有限责任就为负）。[10]

$$E_i^-(t,\theta(t)) \leqslant 0 \tag{6}$$

或者

$$V_i^-(t,\theta(t)) \leqslant \sum_\tau B_i(t-1,\tau,\theta(t-1)) \cdot p(t,\tau,\theta(t)) \tag{6'}$$

在将来某一时期、某一自然状态下，公司的资产价值为零的事实并不意味着公司资产今天的价值也是零；如果有可能公司不会走向破产，那么显然公司现在的价值是正的。但是这确实意味着公司所发行的，在公司有可能破产那一期或者之后到期的债券是风险债券，换句话说，这些债券的到期价值是不确定的，从而，当然这些债券的价格与那些到期价值确定的债券价格是不相同的。所以，一个能够影响公司破产概率（影响公司不能履行其偿债义务的概率）的公司财务政策的变化，将会影响公司所发行债券的价格，而在下面部分的分析中，正是债券价格的不变性起重要的作用。

10　显然，当公司的股票价值为零时，公司当然不能再通过发行新股来筹集偿债资金。

我们将在下面的分析中假设不存在破产的可能性。

2. 家庭

令 $\omega^{j-}(t)$ 表示第 j 个人在该期初的财富水平，$\omega^{j+}(t)$ 为第 j 个人在该期末的财富水平。如果 $E_i^{j+}(t)$ 为在该期末该个人所拥有的第 i 家公司资产的股份，而 $B^j(t,\tau)$ 为本期末该个人所拥有的，在第 τ 期到期的债券的量（由于不存在破产，所以任意一家公司所发行的债券都是无差异的），那么

$$\omega^{j+}(t) = \sum_i E_i^{j+}(t) + \sum_{\tau=t+1}^{\infty} p(t,\tau)B^j(t,\tau)$$

所以，根据式（3），我们得到

$$\omega^{j+}(t) = \sum_i \alpha_i^j(t)V_i^+(t) + \sum_{\tau=t+1}^{\infty} p(t,\tau)\left(B^j(t,\tau) - \sum_i \alpha_i^j(t)B_i(t,\tau)\right) \quad (7)$$

这里，$\alpha_i^j(t) = E_i^{j+}(t)/E_i^+(t)$ 为在 t 期末第 j 个人所持有的第 i 家公司的资产股份的比例。

在下一期的开始，他的资产组合的价值为

$$\omega^{j-}(t+1) = \sum_i \alpha_i^j(t)V_i^-(t+1) + \sum_{\tau=t+1}^{\infty} p(t+1,\tau)\left(B^j(t,\tau) - \sum_i \alpha_i^j(t)B_i(t,\tau)\right) \quad (8)$$

最后，因为在这期间他所获的股利为 $\sum_i \alpha_i^j D_i$，如果他的消费为 $c^j(t)$，那么他的期末财富水平和期初财富水平将通过如下方程式联系起来：

$$\omega^{j+}(t) = \omega^{j-}(t) - c^j(t) + \sum_i \alpha_i^j(t-1)D_i(t) \quad (9)$$

所以，将式（4）、式（7）和式（8）代入式（9），我们对所有的 j，τ，有

$$c^j(t) = \sum_i \alpha_i^j(t-1)\{X_i(t) - I_i(t) + V_i^+(t) - \sum_{\tau=t}^{\infty} p(t,\tau)B_i(t-1,\tau)\} + \sum_{\tau=t}^{\infty} p(t,\tau)B^j(t-1,\tau) - \omega^{j+}(t) \quad (10)$$

式（7）和式（10）可以被理解为定义了个人的消费可能集。给定个人在 $t-1$ 期所拥有的股票集为 $\alpha_i^j(t-1)$ 和债券集为 $B^j(t-1,\tau)$，式（10）定义了该个人在 t 期的消费的价值加上 t 期末的财富水平之和。给定 t 期末个人的财富水平，式（7）定义了 t 期可能的股票和债券拥有情况，而这也就定义了消费的价值加上 $t-1$ 期末的财富总水平之和。注意到，这条可能性曲

线并不取决于股利或者留存收益，这一结果的直觉上的原因将在下一部分的定理之后阐述。

3. 一般均衡

市场均衡要求对第 i 家公司的总的所有者权益价值等于第 i 家公司的资产总价值，即

$$\sum_j E_i^{j+}(t) = \sum_j \alpha_i^j(t) E_i^+(t) = E_i^+(t)$$

或者对所有的 t，i

$$\sum_j \alpha_i^j(t) = 1 \tag{11}$$

类似地，对每一种在给定时期到期的债券的需求和供给应该相等，对所有的 t，τ 有：

$$\sum_i B_i(t,\tau) = \sum_j B^j(t,\tau) \tag{12}$$

二、基本定理

我们现在已经为中心定理的提出和证明做好准备了。

定理1：(a) 假设任何公司在任何状态下都不会破产。(b) 存在一个各种期限的完全安全债券的完全的市场（这里我们所谓的完全安全说的是，在到期时债券所获得的支付是确定知道的；而在不同期到期的债券价格可能是高度可变的）。[11] (c) 所有的公司都已经做出了真实投资的决定（对每一个 i 的计划 k 已经给定）。(d) 假设这里存在一个出清所有市场的一般均衡，在该一般均衡处，在每一期的每一种可能的自然状态下，每一种不同的债券都有一个给定的价格，每一家公司都有一个给定的价值和一套给定的财务政策（一个特定的债务—权益比率，特定的收益留存比例，特定的债券到期结构等)。

那么，将存在另外一个一般均衡解，对比原来的均衡状态，在该均衡状态处，某一家（或者某几家公司）改变了其财务政策的某些（或者所有）方面，但是公司的价值和所有不同到期结构的债券价格（在任意时期，任意自然状态下）保持不变，而且投资者们进行了相应的抵冲投资组合调整，即对所有的 t，τ，j

11 从一个消费的观点来看，不能说什么时候到期的债券是“较安全的”，而在什么时候到期的债券就是“较不安全的”；也就是说，不能说长期债券就比短期债券一定要风险更高。见作者的1970年论文。而这里我要证明的是从公司的角度来看，债券的到期结构是无关的。

$$\Delta B^j(t,\tau) = \sum_i \alpha_i^j(t)\Delta B_i(t,\tau) \tag{13}$$

也就是说，投资者所持有的每一种债券的变化，正好是所有公司的该种债券价值变化乘以该投资者所持有该公司的股份比例之和，且对所有的 i，j，t

$$\Delta\alpha_i^j(t) = 0 \tag{14}$$

或者，等价地

$$\Delta E_i^{j+}(t) = \alpha_i^j(t)\Delta E_i^+(t) \tag{14'}$$

每一个投资者与每一家公司总价值成比例地改变其所持有该公司资产的股份。

证明的论述是很简单的。我将证明，如果公司的价值与所有种类的债券价格（在任何时期，任何状态下）都不改变，那么对任何个人所面对的，可获得的消费可能集均没有改变。由于消费可能集没有改变，所以每个人选择的消费路径也是不变的（即同样地随时间的消费路径，这当然是一个或然计划，它取决于在每一期那种自然状态发生）。为了做到这一点，他如式（13）、式（14）所示的那样改变自己的投资组合计划。最后我们证明，如果经济中不同个体最初所采用的投资组合配置计划集合（即在公司改变其财务政策之前）是一个均衡，从而在每一期、每一种自然状态下市场都出清，那么新的投资组合配置计划集合也将构成一个均衡。

证明：（1）消费可能集是保持不变的。考虑任意可行的消费路径，以及与之相联系的资产组合配置。根据式（7），马上可得，如果在一个给定的公司财务政策的变化之后，$w^{j+}(t)$，$p(t,\tau)$ 和 $V_i^+(t)$ 保持不变，那么式（13）、式（14）所描述的资产组合的变化就是可行的；而且根据式（10），如果这些变化实现了，那么 $c^j(t+1)+w^{j+}(t+1)$ 将保持不变。所以，如果 $c^j(t+1)$ 保持不变，那么 $w^{j+}(t+1)$ 也将保持不变。显然，如果公司的价值和债券的价格保持不变，在预期中的公司财务政策的变化发生之前，w^j 的价值是保持不变的。

由上述论述马上可得出，如果在原来情况下一个消费路径是可行的，那么在公司财务政策变化后的情况下，该消费路径也是可行的，反之也成立。

（2）因为消费可能集合没有改变，那么，如果在原来情况下，相对于其他的消费路径消费者严格偏好于一个给定的消费路径 $\{c^{j*}(t,\theta(t))\}$，那么在新的情况下消费者也将严格偏好于该消费路径。所以，如上所示的资产组合的变化不但是可行的，而且还是最优的。

（3）如果在原来情况下，在任何的时间点上（任何的状态下）所有市场都出清，那么在新的情况下也是这样的。由于 $\alpha_i^j(t)$ 保持不变，那么，如果之前 $\sum \alpha_i^j(t) = 1$，则之后也同样成立，而且所有证券市场都出清（上面所示式（11）均得到满足）。对在给定日期到期的债券的需求的变化由下式给出：

$$\sum_j \sum_i \alpha_i^j(t) \Delta B_i(t,\tau) = \sum_i \Delta B_i(t,\tau)$$

即它刚好等于供给的变化，所以，如果在原来情况下供求相等（式（12）被满足），那么，新的情况下仍然如此。

三、关于定理及其证明的评论

这里，我想做四种评论。在本部分第 1 点，我将为该定理提供一个直觉上的解释。在第 2 点，我将说明，在我们的分析中所使用的假设条件，在多大程度上弱于之前的一些证明所需要采用的假设条件。在第 3 点，我将指出证明的局限在哪里，以及这些局限对定理的一般普适性的重要性。在第 4 点，我们讨论能够消除由公司财务政策变化所导致的“无效率”的竞争性力量。

1. 直观解释

这条定理的一个基本结论就是，人们可以完全“抵冲”公司所采取的任何财务政策。让我们先仅从口头上来讨论一下，个人需要什么样的行动来抵消公司所能采取的各种各样的行动。假设公司降低它的收益发放比例。这意味着公司将会有更多的留存收益，所以，如果那两个基本的会计恒等式成立，那么，要么公司将借得更少（甚至借钱给别的公司），要么公司发行更少的新股。为了抵冲在股利上的损失，即为了保持同样的消费路径，消费者将购买更少的新股或者更少的债券。假设公司将发行更少的新股，在一种情况下，公司权益的价值增大由于新股的发行；在另一种情况下，公司权益的价值增大由于留存收益的增大。从股东的角度来看，这两者是完全等价的。所以，公司收益发放比例的变化没有改变公司的债务—权益比率。另一方面，如果公司减少发行的债券量，那将导致公司的债务—权益比率下降。那么个人将以自己的名义借钱。也许可以将这个过程理解为，就好像个人借更多的钱去购买增加了的公司净资产价值（由于两者是完全等价的，所以这只是理解这个过程的一种便利的方式；因为所有的资金都是无差异的，这两

者之间并无真实的联系)。个人所增加的借贷正好完全抵消公司所减少的借贷,所以市场仍然是出清的。如果公司决定发行更多的3年期债券而减少发行5年期的债券,那么个人可以在他们自己的资产组合中进行完全相反的抵冲调整。

2. 关于定理的普适性

(1) 风险类型、阿罗—德布鲁证券、均值方差分析。应当强调的一点是,在我们的证明中,并没有假设存在两家或者以上的在其他方面完全相同的公司;我们的论证过程也并不需要假设风险类型的存在,或者存在与自然状态同样多的证券种类,也不需要假设收益特征可以由均值方差来体现等。其他关于公司的债务—权益比率无关性的更狭隘的定理证明过程所经常采用的假设。

(2) 资本市场的竞争性。我们并没有做关于资本市场竞争性的假设,唯一的假设就是不存在价格歧视,即,某个人(公司)为某债券(或者股票)所支付的价格与其他个人所支付的价格是相同的。但是市场利率——从而公司所支付的利率——可能会受到公司从市场上所募集的资金量的影响。

(3) 消费者的理性。关于消费者行为的唯一一个限制条件就是,给定一个可行的消费路径集合,消费者总是选择同一条消费路径。所以消费者有可能最大化他的贴现期望理论,但是,要得到我们的结果并不需要这样的限制条件。

(4) 对公司的“控制”。即使个人真的关心他在公司中的政治权利(控制力)(如果公司的真实决定取决于股东,他确实拥有这种权利),如果在公司的决策过程中,股东所起的作用就相当于股东所拥有的那部分股份,那么财务政策仍然是无关的,因为在这两种情况下,任何股东的政治权利都是没有改变的。

也许有人会说,更低的资产总值将使得对公司的控制权转移的可能性更大;但是,在不存在破产的假设下,这是不可能的,因为控制这家公司的股东可以以公司资产作为抵押进行贷款;如果是资产规模小的情况,控制公司的股东可以提高作为一个控制者所需要的最低资产要求,而在资产规模大的情况下,他们这么做更是没有问题。[12]

12 关于债务—权益比率、破产和控制权转移之间关系的一个更深入的讨论请见作者1972年的论文。

（5）引起不确定性源泉。并没有对这方面假设的要求。[13]

（6）均衡的多重性。定理1是一条关于市场均衡的定理。它所说的是，存在无穷多个对经济的一般均衡解，所有这些解在其他方面均相同，只有公司的财务政策和每个人分别所持有的债券和股票价值不同（虽然每个人所持有的每一家公司的股份比例是相同的）。对于给定的财务政策集合，当然也可能存在多个一般经济均衡。通常地，为了保证这种均衡的唯一性，通常需要很强的条件。但是我们的定理所能保证的是，如果对一个给定的债务—权益比率集合存在两个（或者三个……）均衡，那么在任意其他的债务—权益比率集合处也存在两个（或者三个……）均衡。该定理并不涉及哪个均衡将被选择这个问题。

（7）不同的预期。定理的证明并不需要假设个人有相同的预期。所需要达成的唯一的预期的共识就是，在任何自然状态下，公司都不会破产（关于这个假设的讨论见下面部分）。

（8）市场出清。上面分析部分所描述的一条特殊的路径就是，一条每一个个人所做的计划都与其他人所做的计划相一致的均衡路径，即它们都是市场出清的。事实上，分析所唯一需要的一件事就是在时期0市场是出清的。当在0期做出关于资产组合—消费决定时，个人必须有关于在将来所有时期和状态下的，关于价格水平和公司价值的预期。当然，这有可能并没有实现，从而在每一个后续的时期预期被不断地修正。对我们的分析很重要的一点在于，这些对预期的修正并不取决于公司的财务政策，而只取决于“真实事件”。

3. 定理的局限性

（1）预期对于公司财务政策的无关性。我们的分析要求，当公司宣布改变未来期的财务政策时，个人的预期保持不变。

如果事实证明这些预期是公司财务政策的函数时，那么事实上公司的财务政策是会影响公司当期的价值的。那么认为公司的财务政策将影响公司的价值的预期，至少在一个很大概的意义上实现了。但是注意到关于均衡路径的论述说明，没有理由认为这些预期会改变。

（2）个人借贷对公司借贷的不完全替代。也许公司财务政策的无关性

---

13 特别是，技术性不确定性与价格不确定性之间的区别，这在戴梦德的分析中是起着重要作用的，但是在我们这里并不需要这样的假设。而且还应当注意到戴梦德的关于他的结论不依赖于不存在破产的假设的论断是错误的。

定理的一个主要不足就在于，个人借贷不是公司借贷的完全替代。有4个主要原因：1）相比公司借贷，个人借贷的利率更高；2）对个人从市场上借贷的金额的限制；3）交易费用；4）特殊税收条款（对资本所得和公司收入中的利息支付的不同处理）。我已经在其他地方详细地讨论了这些局限性（1969，1973）。这里我只想陈述一些我的观点。

第一，个人借贷时所需要支付的较高的名义利率和对其借贷金额的限制，主要体现了个人的较高的破产概率。换句话说，这也是破产给我们的分析所带来的一般问题的一个特殊表现。

第二，上述前三个原因对于个人无差异的财务政策集合进行了限制，但是没有理由认为这些限制很严重。它们也许意味着公司不能采取所有的资产政策，但是个人仍然对一个在较宽范围里的债务—权益比率无差异。例如，如果公司将要降低它的债务—权益比率，我们的分析并不要求个人从市场上借贷以购买公司发行的更多的股份；而只要求个人降低他们对债券的持有。所以只要该个人所持有股份的公司的债务总和足够大，使得个人是一个净放贷人，而不是一个净借入人，那么个人就是无差异的。这就对在个人的资产组合之中的公司的"平均债务—权益比率"（虽然不是对每一家公司的债务—权益比率）给出了一个下界。如果在那些债务—权益比率中，存在一个有限的破产概率，那么这个限制就可能成为一个重要的限制；这就是说，只有当与破产限制联合在一起，这个限制才变得显著。

第三，这里并不存在对某一家特定公司所采取的债务—权益比率政策做出限制，而只是对这些公司所能够采取的债务—权益比率的政策集合做出限制；也就是说，即使该限制条件是严格的，通常如果有另一家公司降低其债务—权益比率，那么该公司就可以增大其债务—权益比率。这里并不提及最优的债务—权益比率或者留存收益比。

（3）*破产*。在我看来，这里最强的假设就是不存在破产这个假设。

细心的读者也许现在开始纳闷了，我们的证明在什么地方用到了这个假设。根据有限责任法律，显然有（如我前面所提到的）$E_i \geqslant 0$。如果公司发行足够多数量的债券使得在某一期的某种自然状态下 $\sum pB_i > V_i$ 对于与原来（对照）情况下相同的 V_i，那么 E_i 应该是负的，但是这是不可能的。

这个假设并不仅仅对于我们的证明是重要的，而且我将要证明，它对于该定理的合理性也是很重要的。从一个方面来看这个问题，假设在到期时，存在一个正的破产概率的债券的价格与完全安全的债券价格相同是不合理

的。也许有人会说债券名义利率的降低将由一个资产价值的增量所补充，而且在一定的环境下——有和自然状态一样多的证券种类，或者在同质期望的均值方差模型中——确实是这样的。但是在更一般的情况下，破产改变了一个特定个人所面临的机会集合，从而公司的价值也就改变了。不但财务政策是重要的，而且公司的财务与真实决定之间的分离是不可能的（见作者1974年的论文）。[14]

4. 消除“财务管理”中资源“浪费”的竞争性力量

也许有人会质疑，既然财务政策那么不重要，为什么公司还在“资金管理”上浪费资源呢？难道竞争性力量不会迫使所有的公司都忽略财务政策吗？既然担心财务管理的成本，而且它又不能增加公司的价值，那么显然那些在财务管理方面花费资源的公司分发给股东的利润就会要少于那些不在财务管理方面花费资源的公司。关于这个问题有5个答案：

（1）作者忽略了一些重要的因素，特别是税收因素，它确实能使对财务结构的考量带来经济利益。但是这就意味着我相信在不存在税收的情况下，财务管理会淡出历史舞台吗？并不必然，或者如下面几点所说明的，即使是，也是缓慢地。

（2）我已经说明了，如果人们都相信财务政策影响公司的价值，那么它就确实能够影响公司的价值，而且那些忽略这种流行的“偏见”的公司的业绩，可能要差于那些考虑这种因素的公司的业绩。在郁金香球茎热中，可能并没有什么很理性的原因导致郁金香球茎价格的上升，但是只要价格在上涨，至少在短期内，人们可以通过投资于郁金香球茎来赚取“利润”（见下文）。

（3）当然，公司的财务政策和对公司盈利前景的预期之间的关系，可能并不是如上面所分析的那样“不理性”。公司财务政策的变化可能就是关于公司的真正“前景”的一个重要信号。但是，这在我的模型中是不存在

14 这些评论是用来说明本文中的定理，在含义和证明过程方面，与莫迪格利安尼—米勒定理的区别。他们认为财务政策也是无关的。在他们的证明中莫迪格利安尼—米勒利用了风险种类，这样做似乎暗示着可能结果的客观而不是主观概率分布。在他们的证明过程中，保证债务—权益比率对公司的价值不起作用的机制，是个人在同一风险种类中的不同公司之间的套利行为。这种套利行为在我的证明过程中并不起作用。更进一步地，他们的结论是以部分均衡而不是一般均衡为基础的。而且在他们的分析证明过程中，我们并不清楚他们得到的定理是否只对竞争性市场成立，我们也不知道公司的破产概率如何影响他们的结论。但是莫迪格利安尼—米勒分析的基本观点——个人杠杆可以作为公司借贷的替代——在我们这里也起着基础性作用。

这种可能性的，因为在这里根本就没有流动性危机这种事实存在，但是在现实世界中破产可能是非常重要的一个因素，而且银行或者其他金融机构不愿意借钱给公司（例如，迫使公司降低股息以满足公司的流动性要求）的事实，可能说明，那些相对于不太知情的股东们，对公司的真实情况掌握更多信息的人们对公司的前景并不很乐观。[15]

（4）进一步地，没有理由保证在短期内人们对公司的价值的不同评估会导致任何的不一致，或者更一般地，会存在一种力量导致人们重新形成他们对公司的预期，以使得公司价值与公司的财务政策无关。即使我们有两家在任何真实方面都相同的公司（那就是说，在莫迪格利安尼—米勒的体系中，这两家公司属于同一风险种类），并不一定存在一种人们可以（在短期或者中期[16]的）套利的方法[17]。

（5）最后，我们注意到那些浪费在财务管理上的资源可能是相对较少

---

15 实际上，有人可能要说财务政策的信号作用正是其重要功能之一。如果公司从来不派发股利，仅简单地留存所有收益（即使以购买债券的形式），那么很有可能公司可以让股东们比现在更晚才知道公司处在“艰难的状况中”这个事实。这可能可以部分地解释，为什么不管不发股利在避税方面的优势有多大，很多公司仍然继续发放股利。

16 之所以要加这个条件限制是因为，在确定性环境中，可以证明，如果拥有不同价值评估和相同回报的公司采用不同的财务政策的话，那么在有限期内，相对的价值评估必定为无穷。但是在这种环境下，有限可能是更符合实际的。这种价值评估的差异（至少在数学上）与股市中的价格波动经常体现出来的投机性繁荣（或者萧条）是非常相似的。关于这些问题在一个稍微不同的环境中的一般讨论请见卡尔·谢尔（Karl Shell）与作者的著作。

17 为了在一个极端的情况下说明这一点，我们只需要考虑当利润减去投资能被确切地知道，且公司不发新股时的情况。那么此时公司 1 的股利可以被写成（连续时间形式）：

$$D_1(t) = X_1(t) - I_1(t) - rB_1(t) + \dot{B}_1$$

这里所有的债券都被假设为短期债券，赚取一个瞬时收益率 $r(t)$。假设两家除了有一家根本不发行债券的区别外，另外完全一样的公司，这就是说：$D_1(t) = X_1(t) - I_1(t)$。持有其中任意一家公司股票的总收益正是股利和资本所得之和，这里我们要求这两家公司的股票总收益相等。扔掉 X，I 的脚标，我们有：$\rho = \frac{\dot{E}_1}{E_1} + \frac{D_1}{E_1} = \frac{\dot{V}_1}{E_1} + \frac{X-I}{E_1} - \frac{rB_1}{E_1} = \frac{\dot{V}_2}{V_2} + \frac{X-I}{V_2}$ 或者 $\frac{\dot{V}_1}{V_1} - \frac{\dot{V}_2}{V_2} = (r-\rho)\frac{B_1}{V_1} - \frac{(X-I)}{V_1} + \frac{(X-I)}{V_2}$

显然，如果初始时 $\rho = r$，且 $V_1 = V_2$，那么这两家公司的价值永远相同。但是，注意到第二个方程在 $\rho = r$ 和 $V_1 \neq V_2$ 时也能得到满足，在这种情况下将存在市场价值比率的累计性变化。最终这有可能会导致其中一家公司具有一个不同寻常地高，或者不同寻常地低的收益—价值比，而这可能最终会导致一家公司价值的重新评估。但是，正如我们在其他地方详细说明的那样，（在不存在完美的期货市场的情况下）没有理由认为这个过程不会持续很长一段时期。

的（比如，相对于公司的总利润而言），从而消除这种无效率的“竞争性力量”的作用可能是相当有限的。

四、关于任何一家特定公司的财务政策的无关性

上述定理建立了对整个经济而言的财务结构的无关性。我们所采用的一个重要的假设就是不存在破产。我们可以放松这条假设而证明一条关于任何一家特定公司的财务政策无关性的更弱的定理。

定理2：假设经济中存在一个一般均衡，这个一般均衡有如下特征：一个给定的（关于安全债券）的市场利率、一个给定的关于每一家面对破产可能性的公司的风险债券的名义利率、每一家公司拥有给定的市场价值和给定的财务政策（股利—留存收益比，债券到期结构等），而且公司股份的给定比例由第 i 个人持有。

让任意一家公司（或者任意一组公司）改变其财务政策。如果金融中介产品可以无成本地建立，那么对整个经济存在一个新的、拥有相同市场利率的一般均衡，在该新的一般均衡中，每一家公司具有与之前完全相同的市场价值，每一家公司被第 i 个人所拥有的股份比例，或者直接拥有，或者通过金融中介间接地拥有，也与之前完全相同。

既然关于债务—权益比率变化的论证方式与关于其他财务政策变化的论证方式完全类似，我将集中讨论债务—权益比率的变化。假设，在初始均衡中不存在对给定公司股票或者债券的通过金融中介的间接购买。公司改变其债务—权益比率。那么一个金融中介产品被创造出来，用以重构该公司，也就是，购买公司所有的债券和股票，然后，以与之前情况完全相同的比率发行债券和股票。个人所面临的选择集合完全没有改变，因此公司的市场价值、利率等均完全相同。[18]

但是，也许有人要说，个人所面临的选择集合已经改变了，因为，原则上说，个人既可以直接也可以间接通过金融中介来购买公司的债券和股票。所以，如果存在破产的概率，那么，他现在的选择集合要大于之前的选择集

18　更一般地，假设公司的初始债务—权益比率为 d，而且第 k 个金融中介购买该公司 α^k 比例的债券和股票，然后以 d^k 的债务—权益比率发行债券和股票。所以 $1-\sum\alpha^k$ 为公司被个人直接购买（不是通过金融中介）的债券和股票的比例。现在假设公司将债务—权益比率改变为 d'。那么除了 $d^k=d'$ 那家金融中介其他所有的金融中介都不受影响。该家金融中介现在发行债务—权益比率为 d，且购买公司的 $1-\sum\alpha^k$ 比例的债券和股票。

合。这就可能导致对一家给定公司的证券（包括债券和股票）的需求的增加，使得该新的情况并不成为一个均衡。这将存在一个新的均衡情况，在该新的均衡中，该公司的债券和股票比在之前情况中拥有更高的价值。但是，这将意味着之前的情况不可能是一个均衡。因为对一家金融中介而言，它本可以购买和发行同该公司同样比例的债券和股票，从而在任意自然状态下都能够获得公司收益的给定比例，并且能够以在"新的"情况下的债务—权益比率发行债券和股票。这样一来，金融中介的组织者就能够为它们自己创造一个净利润，其价值等于在两种情况下公司价值的差异。

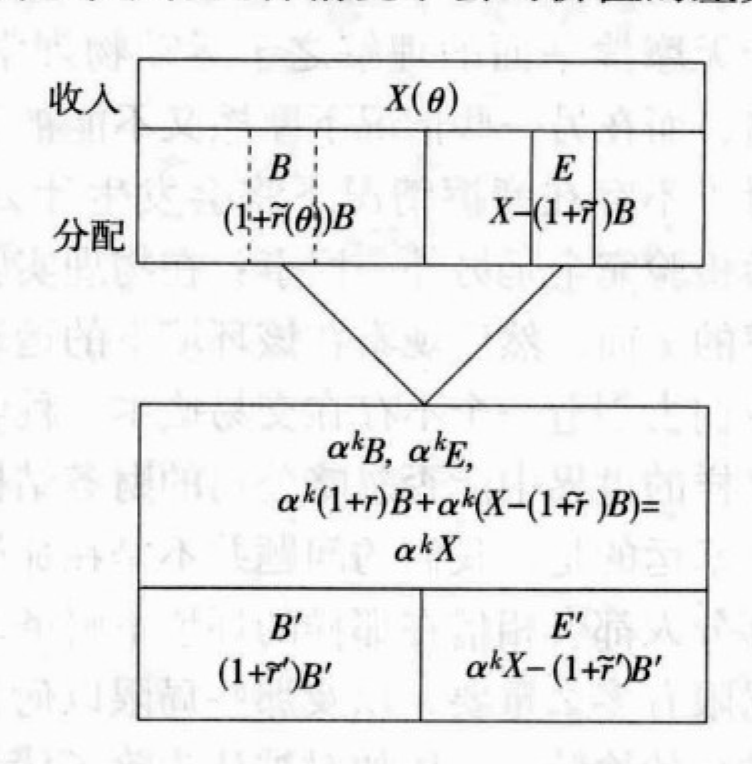


图3　运用金融媒介（$\tilde{r}(\theta)$ = 状态 θ 下对债券的利息支付）

这段论述的重点在于：如果对应于其他公司的一个给定的真实财务决策集合，该公司的财务决策确实起作用的话，那么金融中介的自由进入将使得一系列能够最大化公司价值的证券被发行，而不管公司的债务—权益比率是多少。

如果有人认真地对待金融中介无成本建立这个假设，那么没有理由假设金融中介的扩张过程会在证券种类数等于自然状态数之前停止，当证券种类数和自然状态数相等时，不但每一个特定公司的财务结构是无关的，而且整个经济的财务结构也是无关的。[19]

但是，实际中我们并没有看到金融中介的这种扩张过程的事实说明，要么它们并不必要（这就是说，帕累托最优可以通过有限种类的金融中介来达到，[20]或者定理1的条件得到满足，从而财务结构无关紧要），要么存在相

19　那就是说，所有提供的证券种类数等于自然状态数的财务结构都是等价的。

20　正如在资产组合分离定理成立的环境中那样。见 David Cass and Stiglitz。

对于建立金融中介收益的足够高的交易成本。[21]

五、结论

到现在为止，我们在一些相当一般的条件下，建立了财务政策的无关性定理。关于我们的结果需要指出三种局限性：[22]1）关于公司真实回报的预期取决于公司的财务政策；2）个人借贷并非公司借贷的完美替代；3）破产。在实际应用中，这些局限性是否重要是一个有争议的问题。但是不管它们重要与否，都不能降低该定理在理论上的重要性：对于该定理的理解之于公司财务，就相当于对于无摩擦平面的理解之于运动物理学。在一些实际问题中，摩擦可以被忽略；而在另一些情况下摩擦又不能被忽略。但是，即使当其不能被忽略时，对于不存在摩擦情况下将会发生什么的理解也是很重要的。对该模型的经验检验完全是另外一回事：在物理实验中，我们可以尝试创造一个近似无摩擦的平面，然后观察在该环境下的运动；而在这里要想做类似的事情，需要我们去创造一个不存在交易成本、税收扭曲和其他摩擦的世界，然后观察在这样的世界中能否忽略公司的财务结构。当然，在现实中简直不可能那么做。幸运的是，我们的问题并不是在那样的环境中财务政策是否无关——绝大部分人都会相信在那样的环境下财务政策确实无关——而是：在现实中那些局限有多么重要，以及那些局限以何种方式影响着公司的财务政策。迄今所进行的检验——比如对被认为除了债务—权益比率之外其他方面完全相同的公司（单位规模）的价值的检验——并没有对定理1成立的环境，定理2成立但是定理1不成立的环境，或者定理1和定理2均不成立的环境（在这样的环境中财务政策是很重要的，并且在这样的环境下追求市值最大化的公司会选择能够最大化公司市值的财务政策集合）做区分。我们所需要的是，对那些局限的含义的更深刻的理解，以及对不同理论假说的更精细的检验。

21　包含在“交易成本”中的是获得关于不同种类证券的信息的成本。如果存在一个有限的破产概率，那么债券的购买者就必须要考虑债券的风险。所以破产不但在其确实发生的时候会产生交易成本，而且破产的潜在可能性在债券出售时也产生了交易成本。

22　除了税收的明显的扭曲效应。

参考文献

W. Baumol and B. G. Malkiel, "The Firm's Optimal Debt-Equity Combination and the Cost of Capital," *Quart. J. Econ.*, Nov. 1967, *81*, 547–78.

D. Cass and J. E. Stiglitz, "The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds," *J. Econ. Theory*, June 1970, *2*, 122–60.

P. Diamond, "The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty," *Amer. Econ. Rev.*, Sept. 1967, *57*, 759–76.

N. Kaldor, "Marginal Productivity and the Marco-Economic Theories of Distribution: Comment on Samuelson and Modigliani," *Rev. Econ. Stud.*, Oct. 1966, *33*, 309–19.

J. R. Meyer and E. Kuh, *The Investment Decision: An Empirical Study*, Cambridge 1957.

F. Modigliani and M. H. Miller, "The Cost of Capital, Corporation Finance, and the Theory of Investment," *Amer. Econ. Rev.*, June 1958, *48*, 261–97.

K. Shell and J. E. Stiglitz, "The Allocation of Investment in a Dynamic Economy," *Quart. J. Econ.*, Nov. 1967, *81*, 592–609.

J. E. Stiglitz, "A Re-Examination of the Modigliani-Miller Theorem," *Amer. Econ. Rev.*, Dec. 1969, *59*, 784–93.

———, "A Consumption Oriented Theory of the Demand for Financial Assets and the Term Structure of Interest Rates," *Rev. Econ. Stud.*, July 1970, *37*, 321–51.

———, "On Some Aspects of the Pure Theory of Corporate Finance, Bankruptcies and Takeovers," *Bell J. Econ.*, Autumn 1972, *3*, 458–82.

———, "Taxation, Corporate Financial Policy, and the Cost of Capital," *J. Publ. Econ.*, Jan. 1973, *2*, 1–34.

有风险的市场均衡

论股票市场均衡的无效性*

一、引言

这篇文章证明当不存在完全的市场集合，但是存在一种以上商品时股票市场的均衡，一般而言不是受约束的帕累托最优均衡。整个经济体将会缺乏交换和生产的效率。本文还推导出了要使经济对所有的技术成为一个约束帕累托最优所要满足的必要条件；如果所有人都有相同的、齐次的无差异曲线，那么要么存在单一价格弹性（从而就不存在有效风险），要么所有人都有相同的风险厌恶程度（从而股票市场上也就不存在交易）。

本文是一系列尝试探讨在不确定性条件下经济分配资源有效性的文章之一。阿罗的经典论文（1964）建立了要使市场均衡成为帕累托最优所要满足的一系列充分条件。他要求存在一个完全的风险市场集合，这个条件显然在现实的任何经济中都是无法得到满足的。卡尔·波奇（Karl Borch，1962）

---

* “The Inefficiency of the Stock Market Equilibrium”, The Review of Economic Studies, Vol. 49, No. 2 (Apr., 1982), 241 – 261. 本文初稿收于1979年9月，定稿于1981年9月（Eds.）。

本文是为了在挪威卑尔根（Bergen）市举行的为纪念卡尔·波奇的大会准备的。这是“竞争性股票市场的无效性及其对自然资源损耗的含义”（Stanford，mimeo）一文第一部分的一个简化和修订版本。我要感谢詹姆斯·莫里斯（James Mirrlees）、山迪·格罗斯曼（Sandy Grossman）、巴里·纳尔巴菲（Barry Nalebuff）、罗宾·林德森（Robin Lindsey）、奥利弗·哈特（Oliver Hart），帕萨·达斯古普塔（Partha Dasgupta），维贾·克里希纳（Vijay Krishna），以及各位审稿人的有益建议。同时我也要感谢国家科学基金对本文的研究提供了资金上的大力支持。这篇文章是我在哥伦比亚大学经济系和社会科学中心做访问学者时完成的，同时我也要感谢他们对我的友好接待。

证明当不存在一个完全的市场集合时，市场均衡将不再是帕累托最优。这些结果导致了人们对一些更弱条件集合的寻找，以期在这些更弱的条件下市场均衡能够满足某种更弱意义上的最优。

在观念上，人们可能希望有一个解释市场不完全的理论。有一些理论解释了交易成本和不完全信息。对市场经济的评价应当通过把市场经济与面临相同交易技术与相同信息问题的社会主义经济进行比较来完成。在这样一个比较中，我们也许会发现市场经济在运行中的市场集合方面是不完善的。本文的讨论就能被看作是关于市场经济的一个"微型"的测试：给定运行中的市场集合，它是否能够对资源进行有效的分配；如果是，这仍然不意味着市场经济应当被视做有效的，但是如果不是，显然这说明市场经济是无效率的。

这仍然向我们提出了关于"受约束有效性"的概念问题。解决这个问题的一种方式就是考虑：如果我们将通过生产决策以最大化公司的市场价值的现行的公司经理层，替换为以其他一些标准做出生产经营决策的"公共经理层"，是否有可能在这种新的资源配置情况下每个人的境遇都变好，假设公司的利润仍然按股东持有公司股份的比例分配给股东（而且没有新的风险市场投入运行）。另外，我们也可以考虑，是否有可能通过将市场上这些企业的股份比例重新在该经济中分配（而不是由自由贸易来决定这些股权在经济中的分配）以提高每一个人的福利状况？最后，我们也可以问，在该市场经济中是否存在对市场的"简单的"干预——例如，"一揽子"征税或者补贴计划——以使经济中的一些成员福利水平提高，而不会使经济中其他成员福利水平下降？但是，这仍然可能只被视为关于市场经济的不公平检验：如果这些税收—补贴是依状态而定的，那么政府就可以有效地为经济引进另外一种（或者一系列）证券种类。为了避免这种批评，我们期望将政府的"手"绑得更紧些：我们要求所有的税收或者补贴都不是依状态而定的。这里我们想要强调的是，我们想要展示的是关于市场经济有效性的微型检验，即使该检验被通过，但是这里仍然可能存在政府干预的理论基础，要么在采取措施以影响运行中的市场集合中，要么在引进依状态而定的税收或者补贴中。[1]

在一个只有一种商品的简单股票市场经济情况下，我们所展示的检验等价于戴梦德（1967）所给出的受约束帕累托最优概念。在一个存在股

1 即使关于市场有效性的最小检验，也有可能是误导性的：首先，在不存在风险市场的条件下，检验所推导的原因可能使假设的税收和补贴不可行或者极其昂贵。但是，这个条件在我们这里所检验的特殊环境下，似乎并没有多大作用。在不失一般性的条件下，我们也可以假设它们是依状态而决定的，但如果那样的话，它们将线性依赖于经济中不同公司的产出水平。

票市场的经济中，每一个人都购买每家公司的一部分股票。另外，个人也可以以安全利润率借入或者贷出。戴梦德考察了这样一个经济，在该经济中这些是唯一允许存在的证券种类，并且与市场配置中公司追求最大化的股票价值相反，在该经济中政府支配所有的资源，但是在分配整个经济产出方面受到严格限制：每一个人的分配所得是每一家公司产出的线性函数。随后他证明了：（a）如果每一家企业的生产技术体现随机的齐次性（从而在任意两种状态下企业产出之比与企业规模无关）；（b）每一家企业相信其市场价值与其规模成比例，那么此时市场均衡将是一个约束帕累托均衡。

本文的目的就在于证明，即使在戴梦德所假设的关于市场经济的极端良好的条件下，当存在两种或者以上的产出时，股票市场对于资源的配置结果也不是受约束的帕累托最优。应当注意到，我们这里所设计的模型与金融文献中时常采用的“扩展生成”（Spanning）假设与要素模型是一致的。所以，我们的分析表明，被大家所广为接受的在这些模型中市场均衡是有效的观点，一般而言只对只有一种商品的特殊情况是成立的。

对这个结论有许多直觉上的解释。任何对投资分配或者对不同资产的所有权比例的变化，一般来说都会改变价格的分布。所以风险资产市场化的本质就发生了改变。既然只存在不完全的风险市场，那么可获得风险资产本质的变化将带来福利效应，这是值得考虑的。作为价格接受者的竞争性企业将不会考虑它们的行为对价格分布所带来的影响。不管产出的概率分布如何，对偏好的两个限制性条件可以保证市场是有效的。

这些限制条件保证要么个人根本不关心价格分布，要么所有权或者投资配置的分布不会影响价格的分布。只有当个人具有相同的、齐次的无差异曲线时，所有权的分布才有可能不会对价格分布产生影响，但是只有当不存在交易（因为所有个人都是一样的），或者不存在有效风险（产出的价值为常数，因为所有的个人都具有相同的单一价格弹性需求曲线）时，投资配置的变化对价格分布才不起作用。

我们也可以用另一种方法进行论证。在完全风险市场集合的条件下，我们想让任何个人在任意两种状态下的边际替代率都相等。当不存在完全风险市场集合时，我们不能这么做，但是我们仍然可以有一个风险的更“有效”的分布（平均来说，更接近于边际替代率相等的状态），即如果我们能够改变风险资产的价格分布（从而“利润分布”）。政府能够意识到，它可以通过改变投资配置和对不同资产所有权的比例来改变价格分布。但是市场忽略这种作用。

我们还可以用第三种方法进行分析。在单一产出的条件下，一家企业和一个行业的随机齐次性没有差别：在任意两种自然状态下的产出或者利润之比为常数。而在超过一种产出的条件下，如果某家企业扩大其规模，它想在任意两种状态下的利润之比仍然为常数；但是当一个行业扩大其规模时，在不同自然状态下该行业产出的价格变化可能会有所不同，从而当行业规模变化时，在两个不同状态下产业的利润之比也会发生变化；私人部门具有随机齐次性——将每一种自然状态下价格视做给定——但是社会部门不具有随机齐次性。前者能保证企业愿意追求股票市值最大化，[2,3]也使得它们的假设市值与规模成比例具有合理性。而后者对于保证市场机制的约束最优性是必要的。

从一个更正式的层面来说，竞争性均衡和受约束最优的区别，来自于价格进入关于利润分布的限制性条件（即不管自然状态如何，个体获得公司利润的给定比例）的事实；当一个社会计划者改变所有权配置和投资类型时，它会考虑这些决定将如何影响价格；而在一个竞争性均衡中，这种“价格分布效应”从来都是被企业所忽略的。即使扩展市场的集合——引入期货市场——也得不到（约束）最优。[4]

虽然斯蒂格利茨（1975a）和哈特（Hart，1975）早些时候就注意到了：当存在多于一种的产出时，股市资源配置的非最优性，但是他们那时关于股票市场非有效性本质的特征的分析，并没有本文分析得全面。[5]

---

2 虽然这里我们并没有这么做，但是容易得到（例如，根据 Grossman-Stiglitz，1980 的分析）如果所有的公司都相信，它们的行为对价格的概率分布没有影响的话，那么所有的公司都将面临随机技术齐次性（任意两种状态下的产出比与公司规模无关），从而公司所面临的唯一决定就是关于规模的，从而所有的股东都会希望公司最大化其市场价值。

3 我们忽略了关于破产的常见问题。见 Stiglitz（1972b，1975b）以及 Grossman-Stiglitz（1980）。

4 关于期货市场的问题，在戴梦德的分析中并不存在，因为在那里他假设只有一种商品。

5 哈特（1975）给出了一个多重均衡的例子，在该情形下，其中一个均衡下所有个人的境况都要好于其他均衡下个人的境况。这是一个我在别处称为“结构无效性”（Structural inefficiency）的具体例子（Stiglitz，1972a）。其中一个纳什均衡是帕累托最优，但是没办法保证这就是将要发生的那个均衡。

另一方面，我们这里所讨论的无效性（而且这也是 Stiglitz（1975a）所重点关注的）是边际的无效性；在边际上，私人市场做出了不正确的投资决策。即使这里只有一个唯一的均衡，也是如此。哈特在他（1975）的论文里也做了相似的区分，在那里他建立了一个叫做“强最优”的概念，这个最优就蕴涵了边际无效率的思想。特别地，即使某个均衡是唯一的，也有可能是一个严格次优的均衡。

我们将尝试对竞争性市场均衡成为约束帕累托最优给出一些必要和充分条件集合。虽然我们并没有成功地将该均衡的本质特征做出一个完全的描述，但是我们的基本结果说明存在一个很强的结论，那就是市场竞争均衡不是受约束的帕累托最优。只有当该经济的结构使得风险本身不再起作用，我们才能肯定经济是有效的。

下文第二部分建立基本模型，并列出了股票市场均衡为最优的两个充分条件。第三和第四部分列出了股票市场最优的必要条件。第五部分将我们的分析结论和其他的关于股票市场均衡最优性的结论联系起来。

二、模型

1. 市场均衡

关于股票市场资源配置非最优性的最简单模型，就是一个包含两种由下标 1 和 2 表示的商品的模型。我们设第一种商品为本位商品，而且把两种商品的价格之比计为 p。第 j 个个体的效用为该相对价格 p，和其收入 Y^j 的函数：

$$v^j = V^j(p, Y^j) \tag{1}$$

（V 为间接效用函数）。该个体下一期的收入（财富）取决于当期他如何配置其资产组合。为简单起见，我们假设所有的技术风险都只在第 2 产业中发生，而且包含多重不确定性和规模报酬不变的特点，即，

$$X_2 = g(\theta)I_2 \tag{2}$$

这里 X_2 为第 2 产业在自然状态 θ 下，投资水平为 I_2 时的产出（从该情形扩展到包含多重不确定性且规模报酬递减的情形是很直接的）。我们可以令 $g(\theta) \equiv \theta$（这就简化了符号）。第 1 产业被假设为完全安全：

$$X_1 = (1 + r)I_1 \tag{3}$$

所以如果第 j 个体具有初始财富 I^j，而且他将其财富的 α^j 比例投资于第 1 产业而将剩下的投资于第 2 产业。

$$Y^j = [\alpha^j(1 + r) + (1 - \alpha^j)p\theta]I^j \tag{4}$$

在我们的模型中，我们考虑该个体是直接投资于生产还是投资于某家公司，和然后该家公司再投资于生产是没有差别的。这是由于技术（规模报酬不变）和竞争的假设所致。在市场均衡中处在 i 行业中的任何一家公司的价值都必须等于投资。在竞争性的假设下，它将会相信通过扩大投资规模，可以成比例地提高公司的市场价值。只要公司的价值 V_i，超过投资价值 I_i，那么公司原有股东的净财富就会增加。所以投资会一直持续，直到 $V_i = I_i$。

既然所有的公司股东都相信公司的行为将不会影响在任何自然状态下的价格水平，那么股东将只关心净财富，从而将期望公司去追求市场价值最大化（比较上述的[2]）。

所以，让 β^j 表示第 j 个人在行业 2 中的所有权比例，而 γ^j 为他在行业 1 中所占的股份比例（再一次地，由于我们假设，同一产业中的所有公司都是完全相同且回报完全相关，所以我们没有必要对它们进行区分）。那么

$$(1-\alpha^j)I^j \equiv \beta^j I_2, \sum \beta^j = 1 \quad (5a)$$

$$\alpha^j I^j \equiv \gamma^j I_1, \sum \gamma^j = 1 \quad (5b)$$

$$I_2 = \sum (1-\alpha^j)I^j \quad (6a)$$

$$I_1 = \sum \alpha^j I^j \quad (6b)$$

第 j 个人将 p 与 θ 的关系视做给定，不受其行为的影响；所以对第 j 个人而言，最优的资产组合配置由下式给出

$$\max_{\alpha^j} EV^j(p, (\alpha^j(1+r) + (1-\alpha^j)p\theta)I^j) \quad (7)$$

即

$$EV_Y^j((1+r) - p\theta) = 0 \quad (8)$$

这就意味着对所有的 j，k 有，

$$1 + r = \frac{EV_Y^j p\theta}{EV_Y^j} = \frac{EV_Y^k p\theta}{EV_Y^k} \quad (9)$$

风险资产的加权期望平均必须等于无风险资产的回报，这里权重为相对边际效用。

在这个简单的模型中，理性预期条件下市场均衡的特征被式（9）和另一个方程所刻画：个体所预期的 p 和 θ 之间的关系正是实际的市场关系。在市场均衡中，我们要求对商品 2（比方说）的（消费）需求等于供给，即总供给由下式给出：

$$X_2(\theta, I_2) = \theta I_2 \quad (10)$$

第 j 个人的需求正是价格与其收入的函数：

$$C_2^j = C_2^j(p, Y^j) \quad (11)$$

（这里 Y^j 由式（4）给出）（需求曲线可以很容易地从间接效用函数使用罗尔恒等式

$$C_2^j = - V_p^j / V_Y^j \quad (12)$$

给出；我们将在后面用到这个结果）。

总需求由下式给出

$$C_2 \equiv \sum_j C_2^j \tag{13}$$

所以 p 由下式的解给出

$$C_2 = \theta I_2 \tag{14}$$

2. 受约束帕累托最优

现在需要比较这种市场机制的解与受约束的帕累托最优解。我们假设（就像戴梦德所做的那样）政府能够控制：（a）利润在股东中的分配；（b）企业的产出水平。通过这样假设，在给定其他所有个体的期望效用水平条件下最大化，比如，EV^1。所以对于某个 λ^j 的集合，政府最大化有

$$E\sum_{\{\hat{\beta}^j,\hat{\gamma}^j,\hat{I}_2\}} \lambda^j V^j(p,\hat{Y}^j) \tag{15}$$

这里

$$\hat{Y}^j = \hat{\beta}^j \theta p \hat{I}_2 + \hat{\gamma}^j(1+r)(\bar{I} - \hat{I}_2)$$

其中 $\hat{\beta}^j$ 为从企业 2 的产出中分配给第 j 个人的比例，$\sum \hat{\beta}^j = 1$，而且相似地，$\hat{\gamma}^j$ 为从企业 1 的产出中分配给第 j 个人的比例，$\sum \hat{\gamma}^j = 1$，而且 $\bar{I}$ 为总的投资水平，$\bar{I} = \sum I^j$。

这就给出了解的一阶条件：

$$\lambda^j EV_Y^j(1+r) - \lambda^1 EV_Y^1(1+r) + \frac{1}{I_1} EB\frac{dp(\theta)}{d\gamma^j} = 0 \tag{16a}$$

$$\lambda^j EV_Y^j p\theta - \lambda^1 EV_Y^1 p\theta + \frac{1}{I_2} EB\frac{dp(\theta)}{d\beta^j} = 0 \tag{16b}$$

$$E\sum \lambda^j V_Y^j(\hat{\beta}^j\theta p - \hat{\gamma}^j(1+r)) + EB\frac{dp(\theta)}{dI_2} = 0 \tag{16c}$$

其中

$$\begin{aligned} B &= \sum \lambda^j(V_p^j + V_Y^j\hat{\beta}^j I_2\theta) \\ &= \sum \lambda^j V_Y^j(\hat{\beta}^j\theta I_2 - C_2^j) \end{aligned} \tag{17}$$

（使用罗尔恒等式），且根据式（14），

$$\frac{1}{I_1}\frac{dp(\theta)}{d\gamma^j} = -\frac{\left(\frac{\partial C_2^j}{\partial Y} - \frac{\partial C_2^1}{\partial Y}\right)(1+r)}{\sum_k\left(\frac{\partial C_2^k}{\partial p} + \frac{\partial C_2^k}{\partial Y}\hat{\beta}^k\theta I_2\right)} \tag{18a}$$

$$\frac{1}{I_2}\frac{dp(\theta)}{d\beta^j} = -\frac{\left(\frac{\partial C_2^j}{\partial Y}-\frac{\partial C_2^1}{\partial Y}\right)p\theta}{\sum_k\left(\frac{\partial C_2^k}{\partial p}+\frac{\partial C_2^k}{\partial Y}\beta^k\theta I_2\right)} \tag{18b}$$

和

$$\frac{dp(\theta)}{dI_2} = -\frac{\sum_k\frac{\partial C_2^k}{\partial Y}(\beta^k p\theta-\gamma^k(1+r))}{\sum_k\left(\frac{\partial C_2^k}{\partial p}+\frac{\partial C_2^k}{\partial Y}\beta^k\theta I_2\right)} \tag{18c}$$

式（16）的重点在于，对于风险公司投资配置、收入分配以及所有者权益配置对价格分布的影响，都是我们需要考虑的因素。既然在竞争性市场条件下，对价格分布的影响（根据假设）被忽略，那么一般来说，竞争性市场配置都不是（受约束的）帕累托最优。

（1）交换有效性。条件式（16a）和式（16b）可以被看成是在给定投资配置情况下，实现交换有效性的条件。用式（16a）除式（16b），我们得到（根据式（18a）和式（18b））

$$\frac{EV_Y^1p\theta}{EV_Y^1(1+r)} = \frac{\lambda^jEV_Y^jp\theta+EA^jBp\theta}{\lambda^jEV_Y^j(1+r)+EA^jB(1+r)} \tag{19}$$

其中

$$A^j = \frac{-\left(\frac{\partial C_2^j}{\partial Y}-\frac{\partial C_2^1}{\partial Y}\right)}{\sum_k\frac{\partial C_2^k}{\partial p}+\frac{\partial C_2^k}{\partial Y}\beta^k\theta I_2}$$

对照式（9）和式（19），我们看到要使市场满足交换有效性的一个必要条件为

$$\frac{EA^jBp\theta}{EV_Y^jp\theta} = \frac{EA^jB}{EV_Y^j} \tag{20}$$

很显然，一般而言，式（20）是得不到满足的。实现交换有效性的充分条件为：[6]

（i）$A^j\equiv 0$，即所有个体关于商品 2 的边际消费倾向都相等。对所有的 j 有

6　更精确地说，条件（i）和（ii）是一阶条件式（16a）和式（16b）与市场均衡条件式（9）重合的必要条件。如果对于式（16）存在多重解的话，那么经济仍然有可能不是交换有效的，而且市场均衡要么对应于一个局部（不是全局）最大，要么对应一个局部最小。

$$\frac{\partial C_2^j}{\partial Y} = \frac{\partial C_2^1}{\partial Y} \tag{21}$$

这对于所有价格水平和收入水平均成立。这一点对于无差异曲线的含义已经被广泛地研究过了。如果所有人都具有相同的、齐次的无差异图形，那么，式（21）显然成立。

这个条件保证我们的收入分配对价格没有影响（从而对价格分布也没有影响）。

（ii）$B=0$。为了理解这个定义，我们定义

$$\begin{aligned} T^j &= \hat{\beta}^j \theta I_2 - C_2^j \\ &= (1-\alpha^j) I^j \theta - C_2^j \end{aligned} \tag{22}$$

$\beta^j \theta I_2$ 为个人对于 X_2 的"所有权"（通过他对该行业企业的股份）。所以 T^j 为第 j 个人（隐含的）净贸易额（如果 $T^j > 0$，那么就意味着第 j 个人在出售 X_2，而如果 $T^j < 0$，则意味着他在购买 X_2）。

将式（22）代入式（17），我们得到

$$B \equiv \sum \lambda^j V_Y^j T^j \tag{17'}$$

$V_Y^j T^j$ 衡量的是价格水平的变化对个体 j 的福利水平的影响。如果 $T^j = 0$，那么实际上，该个体既不出售也不购买该风险商品，所以价格水平的变化既不使他的境况变好，也不使他的境况变差。条件 $B=0$ 意思是由价格变化所导致的福利水平的变化的加权平均值（其中权重为收入的社会边际效用）必须等于零，对于任何（相关的）价格水平 p。这是一个极其苛刻的限制条件，这一点我们将在下面的分析中说明。此时，我们注意到 $B=0$，从而关系到经济的交换有效性的两个充分条件为：

（i）所有的个体都是相同的，从而 $T^j=0$；或者

（ii）所有的个体都具有柯布—道格拉斯形式的效用函数，可能有不同的指数；在这种情况下 $p\theta$ 为常数；从而不存在风险，所以经济为受约束的帕累托最优。[7]

（2）**生产有效性**。从式（16a）中我们得到

$$\hat{\gamma}^j \lambda^j E V_Y^j = \hat{\gamma}^j \lambda^1 E V_Y^1 - \hat{\gamma}^j E B A^j$$

对于 j 加总，

7　既然不存在风险，那么 V_Y^j 为常数，且 $\lambda^j V_Y^j = \lambda^1 V_Y^1$，所以

$$B = \lambda^1 V_Y^1 \sum T^j = 0$$

关于 $p\theta$ 为常数的证明方法，是直接的但是却是冗长乏味的，我们将在第4部分给出。

$$\sum_j \hat{\gamma}^j \lambda^j EV_Y^j = \sum_j \hat{\gamma}^j \lambda^1 EV_Y^1 - \sum_j EBA^j \hat{\gamma}^j$$
$$= \lambda^1 EV_Y^1 - \sum_j EBA^j \hat{\gamma}^j$$

类似地

$$\sum_j \hat{\beta}^j \lambda^j EV_Y^j p\theta = \lambda^1 EV_Y^1 p\theta - EB \sum_j A^j \hat{\beta}^j p\theta$$

代入式（16c），我们得到

$$\lambda^1 EV_Y^1 (p\theta - (1+r)) + EB \frac{dp}{dI_2} - EB\left(\sum A^j(\hat{\beta}^j p\theta - \hat{\gamma}_j(1+r))\right) = 0 \tag{16c'}$$

根据式（9），我们看到在市场均衡配置中，

$$EV_Y^1(p\theta - (1+r)) = 0 \tag{9'}$$

式（16c′）连同式（9′）意味着市场配置是受约束的帕累托最优的一个必要条件为：

$$EB\left[\frac{dp}{dI_2} - \sum A^j(\hat{\beta}^j p\theta - \hat{\gamma}^j(1+r))\right] = 0 \tag{23}$$

作为另一种方式，根据式（18c），我们重写式（23），

$$EB\left[\frac{-\dfrac{\partial C_2^1}{\partial Y}(p\theta - (1+r))}{\sum_k\left(\dfrac{\partial C_2^k}{\partial p} + \dfrac{\partial C_2^k}{\partial Y}\hat{\beta}^k \theta I_2\right)}\right] = 0 \tag{23'}$$

本文要论述的重点就在于说明在式（23）（或者式（23′））只有在非常严格的，即使比起那些交换有效性所要求的必要条件，还要严格得多的限制条件下才能够被满足。

如果经济满足交换有效性，那么根据式（20）

$$\hat{\beta}^j EA^j Bp\theta = \hat{\beta}^j \frac{EV_Y^j p\theta}{EV_Y^j} EA^j B$$

再使用式（9），对 j 求和，我们得到

$$E\sum \hat{\beta}^j A^j Bp\theta = (1+r)\sum_j EA^j B\hat{\beta}^j \tag{24}$$

从而式（23）变为

$$EB\left[\frac{dp}{dI_2} - \sum_j A^j(\hat{\beta}^j - \hat{\gamma}^j)(1+r)\right] = 0 \tag{25}$$

由交换有效的条件显然得不到生产有效的结论。特别地，如果 $A^j = 0$（回忆一下，这是一个关于交换有效性的充分条件），我们仍然要求 $EBdp/dI_2 = 0$。当不

存在再分配时的收入效应时，市场机制是受约束帕累托最优的一个必要条件为

$$\Omega \equiv EB\frac{dp}{dI_2} = 0 \tag{26}$$

根据式（17′），我们可以将此式重写为

$$E\sum \lambda^j V_Y^j T^j p\frac{d\ln p}{dI_2} = 0 \tag{26′}$$

显然，正如我们早先所注意到的那样，如果投资的配置对两种商品的相对价格没有影响（$dp/dI_2 \equiv 0$）的话，那么 $\Omega = 0$。但是这种情形意义并不大，因为它要求两种商品为彼此的完全替代品。

即使 I_2 的改变并不导致两种状态下相对利润的改变，即对所有的 θ 有

$$\frac{d\ln p}{dI_2} = 常数$$

一般来说，Ω 也不会为零。另外，我们还要求

$$E\sum \lambda^j V_Y^j p T^j = 0 \tag{26″}$$

即“净出售额”的价值（pT^j）与收入的边际效用之间的（平均）协方差应当为零。

从而，立即可得要使式（26）成立的三个充分条件：

（a） 所有的个体都是一样的，此时，如我们早先所注意到的那样，$T^j \equiv 0$ 且 $B \equiv 0$；

（b） 对所有的个体有 $dV_Y^j/d\theta = 0$，而且市场具有交换有效性，交换有效性保证了

$$\lambda^j EV_Y^j = \lambda^1 EV_Y^1$$

从而

$$\sum_j E\lambda^j V_Y^j T^j \frac{dp}{dI_2} = \lambda^1 V_Y^1 E\sum_j T^j \frac{dp}{dI_2}$$

$$= \lambda^1 V_Y^1 E\frac{dp}{dI_2}\sum T^j = 0 \tag{26‴}$$

在附录中，我们发现条件 $dV_Y^j/d\theta = 0$，即使比收入中性条件（$V_{YY} = 0$）还要苛刻得多；即使所有个体都具有收入风险中性，要满足 $dV_Y^j/d\theta = 0$，个体对商品的需求还必须与收入无关。

（c） 所有的个体都具有柯布—道格拉斯效用函数，此时，正如我们前面所述的，$p\theta$ 为常数且 $B \equiv 0$。

在这一部分，我们发现：要使市场均衡成为受约束下的帕累托最优条

件，与代表竞争性市场均衡的条件大不相同。一般而言，给定了投资模式，市场将不会以受限制的帕累托最优的方式根据不同资产配置所有者权益；一般而言，这种情况下市场不具有交换有效性的特点。而且市场在不同产业之间对投资额的分配也不是有效的。在两种情况下，无效性都来自于，配置决策会对风险资产的价格分布产生影响这样一个事实。正如我们在导言中提到的，就好比我们用一种风险资产替换了另一种风险资产，市场忽略了这种作用，而"一个中央计划者"是不会忽略的。

因此，我们的结果明显地意味着市场机制的（受限）帕累托最优不成立。但是，是否还存在一系列保证市场帕累托最优的假设（比如说，关于个体偏好）就不是那么清楚了。在接下来的部分，我们将证明这类假设正是那些能够保证要么不存在风险，要么即使有风险，股票市场上也不存在交易的假设。

三、偏好相同条件下最优的必要条件

在前面的部分，我们证明了：一般来说市场配置资源的功能既缺乏交换有效性也没有生产有效性。有效性所需要的条件看起来是非常苛刻的。

事实上，我们只能得到三种保证有效性的充分条件，其中一种还是非常没有意义的（也就是所有的人都无差异），另一种条件使得风险市场不再有意义（所有人都具有柯布—道格拉斯形式的效用函数），而最后一种与效用函数的具体形式和经济中的技术并不直接相关（上述条件 b）。我们自然要问的问题是，是否存在更弱一点但同样能保证有效性的条件？或者说，我们是否已经抓住了市场失灵的根源？

本部分和下一部分的目的在于证明：事实上，在很大程度上应当假定市场在资源配置上是无效率的。我们通过如下质疑完成了对上述假设的检验：在什么样的效用函数形式下，不管技术形式和经济中的初始禀赋，市场配置是有效的。[8] 既然，我们主要关心的是市场有效地配置资源的能力，那么我们着重于市场具有交换有效性的情况；我们假设所有个人都具有齐次的无差异曲线。如我们前面所述，这就意味着，所有权益的重新分配不会对价格水平的分布产生影响（而且也正是这个条件保证了市场的交换有效性）。个人仍然可以在对风险的态度方面有所差异，而且也正是这种风险厌恶态度的差

---

8 从我们的分析过程可以清楚地看到，我们也可以选择提供一条具有如下形式的一般的无效性定理："除了在下面提及的条件下，对于任意一个具体的效用函数形式集合，使得该经济成为受约束帕累托最优的禀赋，和 θ 的概率分布集合的测度为零。"

异，才为股票市场的交易提供了基础。

这里我们得到，当所有的个人具有无差异的齐次性的偏好时，市场实现资源配置（对所有的技术形式和禀赋）的受约束的帕累托最优的唯一条件为：

(i) 所有的个人具有相同的、常相对风险厌恶系数，从而在股票市场上不存在交易；

或者

(ii) 所有的个人具有单一的需求价格弹性，从而每个产业的产出价值为给定的；从本质上来说，这时也不存在风险，股票市场上也不会存在交易。

关于偏好的无差异性和齐次性的假设，意味着价格水平不受收入分配的影响。这将大大简化分析。第 j 个人的间接效用函数可以被写成：

$$V^j = \phi^j(Y^j u(p)) \tag{27}$$

$$V_Y^j = \phi^{j\prime} u(p), V_p^j = \phi^{j\prime} Y^j u' \tag{27a}$$

从而，根据罗尔恒等式，

$$C_2^j = -\frac{Y^j u'}{u}$$

令 s^j 为商品 1 所占的支出比例，那么

$$s^j = \frac{C_1^j}{Y^j} = \frac{u'p}{u} + 1 \tag{28}$$

为简便起见，定义

$$v(p) \equiv \frac{1 - s(p)}{s(p)} = -\frac{u'p/u}{1 + u'p/u}$$

在均衡时，

$$v(p) = \frac{1 - s(p)}{s(p)} = \frac{pX_2}{X_1} = \frac{p\theta I_2}{(1 + r)I_1} \tag{29}$$

所以

$$\frac{I_2}{p}\frac{dp}{dI_2} = \frac{v}{v'p - v}\left(1 + \frac{I_2}{I_1}\right) \tag{30a}$$

并且有替代弹性为

$$-\frac{\theta}{p}\frac{dp}{d\theta} = -\frac{v}{(v'p - v)} \equiv \frac{1}{\sigma} \tag{30b}$$

所以

$$\frac{v'p}{v} = (1 - \sigma)$$

将式（30a）和式（30b）代入式（26），再根据式（17′），我们得到：

$$-\Omega = \frac{\bar{I}}{I_2 I_1}\sum E\lambda^j \phi^{j'} u(p) T^j \frac{p}{\sigma} \tag{21′}$$

现在我们考虑两组的情况，分别由下标 a 和 b 表示。那么此时 $T^a = -T^b$。给定政府的转移支付政策为最优的，那么

$$EDu(p) = 0 \tag{31a}$$

这里 D 为（加权）边际效用之差

$$D \equiv \lambda^a \phi^{a'} - \lambda^b \phi^{b'} \tag{31b}$$

（式（31a）正是式（16a）的一个特殊情况。）

所以式（21′）可以被重写为

$$-\Omega \equiv \frac{\bar{I}}{I_1 I_2} E\frac{Du(p) T^a p}{\sigma} = \frac{\bar{I}}{I_1 I_2} E[Du(p) - EDu(p)]\left[\frac{T^a p}{\sigma} - E\frac{T^a p}{\sigma}\right] \tag{32}$$

从上式我们可以得到，对于 θ 所有可能分布的市场机制，均满足最优的一个必要条件为

$$\frac{T^a p}{\sigma}$$

它与 θ 无关，或者 $D \equiv 0$。

（a）$D \equiv 0$ 的条件。对于所有可能的参数值 $D \equiv 0$ 要求（令 $\lambda^a / \lambda^b \equiv \lambda$）

$$\frac{d\lambda}{\lambda} \equiv \rho^a \frac{dY^a}{Y^a} - \rho^b \frac{dY^b}{Y^b} \tag{33}$$

其中 $\rho^j = (-V^j_{YY} Y^j / V^j_Y) = -(\phi^{j''} Y^j u / \phi^{j'})$ = 相对风险厌恶的度量。由于

$$\frac{p\theta I_2}{(1+r) I_1} = v(p) \tag{34}$$

那么

$$p\theta = kv(p)$$

其中，$k = (1+r) I_1 / I_2$。

不是一般性，我们可以将 Y^j 写成

$$Y^j = M^j + N^j v(p) \tag{35}$$

从而

$$d\ln\lambda \equiv \frac{\rho^a (dM^a + v(p) dN^a)}{M^a + v(p) N^a} - \rho^b \frac{dM^b + v(p) dN^b}{M^b + v(p) N^b} \tag{33′}$$

因为

$$\frac{D}{\lambda^b} \equiv \lambda\phi^{a'}[u(p)(M^a + N^a v)] - \phi^{b'}[u(p)(M^b + N^b v)]$$

$dD/dp=0$ 意味着

$$\rho^a\left(\frac{u'p}{u}+\frac{N^a v'p}{M^a+N^a v}\right)=\rho^b\left(\frac{u'p}{u}+\frac{N^b v'p}{M^b+N^b v}\right) \tag{36}$$

或者

$$\rho^a\left(-(1-s)+\frac{N^a}{Y^a}\frac{(1-s)}{s}(1-\sigma)\right)$$

$$=\rho^b\left(-(1-s)+\frac{N^b}{Y^b}\frac{(1-s)}{s}(1-\sigma)\right) \tag{36'}$$

现在考虑对问题的参数值的扰动（给定 λ 保持不变）。很明显，根据式（33′）

$$\rho^a\frac{dM^a+vdN^a}{Y^a}=\rho^b\frac{dM^b+vdN^b}{Y^b} \tag{37}$$

用式（36′）除式（37），我们得到

$$\frac{dM^a+vdN^a}{(1-s)\left[N^a\frac{(1-\sigma)}{s}-Y^a\right]}=\frac{dM^b+vdN^b}{(1-s)\left[N^b\frac{(1-\sigma)}{s}-Y^b\right]} \tag{38}$$

或者

$$dM^a+vdN^a=\frac{N^a\frac{(1-\sigma)}{s}-Y^a}{N^b\frac{(1-\sigma)}{s}-Y^b}(dM^b+vN^b) \tag{38'}$$

式（38′）左手边关于 v 为线性。所以右边也应该关于 v 为线性，即

$$N^a(s-\sigma)-sM^a=\kappa[N^b(s-\sigma)-sM^b]$$

或者

$$\frac{(s-\sigma)}{s}(N^a-\kappa N^b)=M^a-\kappa M^b \tag{39}$$

从而，要么 $(s-\sigma)/s$ 为常数，要么 $N^a-\kappa N^b=M^a-\kappa M^b=0$：

(i) $\frac{s-\sigma}{s}=$ 常数 (40a)

存在一组特殊的效用函数形式可以使上述式子成立，可以通过解下列微分方程得到某些 c 值：

$$\sigma=1-\frac{v'p}{v}=c\cdot\frac{1}{1+v}=c\cdot s \tag{41}$$

然后再解下述微分方程：

$$\frac{d\ln u}{d\ln p} = -\frac{v^*(p)}{1 + v^*(p)'} \tag{42}$$

其中，$v^*(p)$ 为式（41）的解。

在常弹性这一类效用函数中，唯一满足式（41a）的效用函数形式为（平凡地）单一替代弹性，即 $p\theta$ 为常数，从而不存在风险。

到现在为止的分析中，我们都假设关于产出的不确定性只发生在部门2。通过完全相同的分析方法，如果所有的不确定性只发生在部门1，我们可以得到

$$\frac{\sigma}{1 - s} = 常数 \tag{40b}$$

的条件。式（40a）连同式（40b）说明 σ 为常数，从而我们的分析说明，如果对所有可行的问题参数值都有 $D = 0$，对所有的 p 成立，那么个体必须具有单一价格弹性。

（ii）$M^a = \kappa M^b$，$N^a = \kappa N^b$

两类个体具有相同的资产组合，即他们必须具有相同的（常）相对风险厌恶系数，或者说事实上不存在风险（$\sigma = 1$）。

（b）$T^a p/\sigma$ 为常数的条件。类似地，只有在很特殊的条件下，$T^a p/\sigma$ 才能保持为常数。回忆一下式（22）和式（28），

$$\begin{aligned} pT^a &= \beta^a p\theta I_2 - (1 - s(p))Y^a \\ &= \beta^a p\theta I_2\left(1 - \frac{v}{1 + v}\right) - \frac{v\gamma^a}{1 + v} I_1(1 + r) \\ &= [\beta^a - \gamma^a] I_1(1 + r)\frac{v}{1 + v} \end{aligned} \tag{43}$$

对任意的 p，β^a 和 γ^a 值 $T^a p/\sigma$ 都必须保持不变。因此，要么 $\beta^a \equiv \gamma^a$（所有的个体都是无差异的），要么

$$\frac{v}{(1 + v)\sigma} = \frac{1 - s}{\sigma} \equiv 常数 \tag{44}$$

与前面所得到的关于 $D \equiv 0$ 的条件式（40b）完全相同。

因此，我们证明了，如果所有的个体都具有齐次无差异图像的话，那么，如果对于所有的技术，市场机制的配置结果都是一个受限的帕累托最优，则要么所有的个体都是无差异的，要么 $p\theta$ 必须为常数。在两种情况下股票市场都没有起到任何的风险分担或者分散的功能，在头一种情况下，是因为所有的个人都是无差异的，在后一种情况下，是因为 $p\theta$ 为常数，从而该经济中根本就不存在风险。

四、偏好的不同

在前面的部分中，我们证明了如果个人具有相同偏好，而只在他们面对风险的态度方面有所差异的话，那么市场为受约束帕累托最优（对所有可能的技术）的唯一可能的情况，是所有的个人都具有单一的价格弹性。

在这一部分，我们将建立一个相对而言更一般的结果：即使当个人具有不同的偏好时，如果所有的个人具有单一的价格弹性，那么市场仍然是一个受限帕累托最优，但是，原因只是此时不存在市场风险——$p\theta$ 为常数。

在另一方面，我们也说明了这个结果是不稳健的。市场结构的稍微变动——关于价格水平 p 的期货市场的引进——会导致两个重要的结果：

（a）（即使所有的个体仍然具有单一的价格弹性）此时将存在总体的市场风险；而且

（b）市场将不再满足受约束的帕累托最优条件，除了在我们之前已经讨论过的一个非常特殊的情况下：所有的个人都具有相同的无差异曲线和单一的价格弹性。

为了得到比较简单明了的结果，我们假设只存在两个组别。第 j 个人在商品 1 上花费他收入的 s^j，其中 $0<s^j<1$。

从而，我们可以将其效用函数写成一个柯布—道格拉斯形式效用函数的某种凹转换：

$$U^j = \phi^j((C_1^j)^{sj}(C_2^j)^{1-sj}),\phi^{j'}>0,\phi^{j''}<0$$

从而，得到一个如下形式的间接效用函数：

$$V^j = \phi^j(\alpha^j p^{sj-1}Y^j),\text{其中}\alpha^j = (s^j)^{sj}(1-s^j)^{1-sj}$$

（式(27) 的当 $u(p) = p^{s-1}$时的特殊情况）。

由于我们在这部分的分析目的，就是看引入一个额外的市场——对下一期的 X_2 的价格水平的一个赌博（这就叫做一个期货市场）对总体风险的影响，我们令 δ^j 为第 j 组在期货市场上的购买（出售）量，q 为现行价格水平。从而

$$Y^j = \beta^j p\theta I_2 + (1+r)[I^j - \delta^j q - \beta^j I_2] + \delta^j p \tag{45}$$

如果 δ^j 被最优地决定，那么

$$EV_Y^j\frac{p}{q} = EV_Y^j(1+r)$$

或者

$$\frac{EV_Y^j p/q}{EV_Y^j} = 1 + r \tag{46}$$

再一次，回报的加权平均必须与无风险资产的回报相等。

现在在均衡处，商品 1 的需求必须与供给相等，

$$I_1(1+r) = s^aY^a + s^bY^b \tag{47a}$$

且对“期货”的需求与供给也应该相等，即

$$\delta^a = -\delta^b \tag{47b}$$

所以，根据式（45）

$$I_1(1+r) = (s^a\beta^a + s^b\beta^b)p\theta I_2 + (s^a - s^b)\delta^a p + kI_1(1+r) \tag{48}$$

其中

$$k = (s^a\gamma^a + s^b\gamma^b)$$

式（48）有一个重要的含义，那就是 $p\theta$ 为 p 的一个线性函数：

$$p\theta = \frac{I_1(1+r)(1-k) - (s^a - s^b)\delta^a p}{I_2(s^a\beta^a + s^b\beta^b)} = z_1 + z_2 p \tag{49}$$

将式（49）代回到式（45）中，我们看到 Y^j 关于 p 为线性：

$$Y^j = Q^j + R^j p \tag{50}$$

其中

$$Q^j = (1+r)[I^j - \delta^j q - \beta^j I_2] + \frac{\{\beta^j I_1(1+r)(1-k)\}}{s^a\beta^a + s^b\beta^b} \tag{51a}$$

$$R^j = \delta^j - \frac{\beta^j(s^a - s^b)\delta^a}{s^a\beta^a + s^b\beta^b} \tag{51b}$$

对上述事实的证明将是我们分析中非常重要的一点。式（49）具有一个非常重要的含义：在不存在期货市场的情况下（$\delta \equiv 0$），如果所有人都具有单一价格弹性，但是有不同的偏好（花费在两种商品上的支出比例），那么加总的需求曲线将仍然具有单一的价格弹性；这样不会存在市场风险，从而市场配置就是一个受约束的帕累托最优。但是当我们引进了期货市场以后，加总的需求曲线将不再具有单一的价格弹性，市场配置结果将不再是一个受约束的帕累托最优。

为了说明这一结果，我们需要利用受约束帕累托最优的一阶条件和需求函数的特殊性质。特别地，我们注意到（根据式（48））：

$$\frac{d\ln p}{dI_2} = -\frac{\bar{s}p\theta + (1-k)(1+r)}{I_1(1-k)(1+r)} \equiv -\frac{m_1 p\theta + 1}{I_1} \tag{52a}$$

$$\frac{d\ln p}{d\gamma^a} = -\frac{(s^a - s^b)}{1-k} \equiv m_2 \tag{52b}$$

$$\frac{d\ln p}{d\beta^a} = -\frac{(s^a - s^b)p\theta I_2}{I_1(1-k)(1+r)} \tag{52c}$$

$$\frac{d\ln p}{d\delta^a} = -\frac{(s^a - s^b)p}{I_1(1-k)(1+r)} \tag{52d}$$

其中$\bar{s} = s^a\beta^a + s^b\beta^b$。

关于交换有效性的一阶条件，现在可以被改写成：

$$ED\left(1 - \frac{pT^a(s^a - s^b)}{I_1(1-k)(1+r)}\right) = 0 \tag{53a}$$

$$ED\left(1 - \frac{pT^a(s^a - s^b)}{I_1(1-k)(1+r)}\right)p\theta = 0 \tag{53b}$$

$$ED\left(1 - \frac{pT^a(s^a - s^b)}{I_1(1-k)(1+r)}\right)p = 0 \tag{53c}$$

这里

$$T^j \equiv \beta^j\theta I_2 + \delta^j - c_2^j \tag{54}$$

现在注意到pT^a关于价格p为线性的：

$$\begin{aligned} pT^a &= p\beta^a\theta I_2 - (1-s^a)Y^a + p\delta^a \\ &= s^a(\beta^a p\theta I_2 + \delta^a p) - (1-s^a)\gamma^a I_1(1+r) \\ &= s^a\beta^a I_2 z_1 - (1-s^a)\gamma^a I_1(1+r) + s^a p(z_2\beta^a I_2 + \delta^a) \\ &\equiv z'_3 + \frac{s^a s^b \delta^a}{\bar{s}}p \end{aligned} \tag{55}$$

从而，式（53）可以被写为

$$ED(z_3 + z_4 p) = 0, \quad ED(z_3 + z_4 p)p = 0 \tag{56}$$

其中

$$z_3 = -\frac{z'_3(s^a - s^b)}{I_1(1-k)(1+r)} + 1$$

$$z_4 = -\frac{(s^a - s^b)s^a s^b \delta^a}{I_1(1-k)(1+r)\bar{s}} \tag{57}$$

关于生产效率的一阶条件可以被写成

$$E\sum \lambda^j V_Y^j(\beta^j p\theta - \gamma^j(1+r)) + EDpT^a\frac{d\ln p}{dI_2} = 0 \tag{58}$$

在市场均衡中

$$EV_Y^j p\theta = EV_Y^j(1+r) \tag{59}$$

因此，要使一个市场均衡成为一个受约束的帕累托最优，那么有

$$(1+r)ED(\beta^a - \gamma^a) + EDpT^a\frac{d\ln p}{dI_2} = 0 \tag{60}$$

根据组合均衡条件可得

$$\lambda^a EV_Y^a p\theta = \lambda^a EV_Y^a(1+r) \tag{61}$$

$$\lambda^b EV_Y^b p\theta = \lambda^b EV_Y^b(1+r) \tag{62}$$

减去

$$EDp\theta = (1+r)ED \tag{63}$$

再将式（63）和式（52a）代入式（60），我们得到（经过一些化简处理）：

$$ED = 0 \tag{64}$$

但是，同时根据式（49）和式（63）：

$$EDp\theta = z_1 ED + z_2 EDp = (1+r)ED \tag{65}$$

所以，要么

$$z_2 = 0 \tag{66}$$

要么

$$EDp = 0 \tag{67}$$

如果 $z_2=0$，那么 $p\theta$ 为常数：从而不存在市场风险。

根据式（64）和式（57），我们现在可以将关于生产有效性的条件改写为：

$$EDpT^a \frac{d\ln p}{dI_2} = -\frac{s^a s^b \delta^a}{I_1(1-k)(1+r)} EDp^2\theta = 0 \tag{68}$$

这就要求[9]要么

---

9 如果 $p\theta$ 不是常数（$z_2 \neq 0$），且 $D \neq 0$，那么我们可以找到这样的概率分布，使得即使在竞争性配置中，$ED=0$，$EDp=0$，$EDp^2\theta = z_2 EDp^2 \neq 0$。为了简单地说明这一点，我们注意到竞争性的配置要求

$$EV_Y^i\ (p\theta - (1+r)) = 0, \quad i=a,\ b \tag{71}$$

$$\frac{EV_Y^a p}{EV_Y^a} = \frac{EV_Y^b p}{EV_Y^b} \equiv q \tag{72}$$

式（64）连同式（67）说明式（72）成立。让我们先给定 $[I_1, I_2, \gamma^a, \beta^b, \delta^a]$，从而对每一个 θ 给定了 Y^a 和 p。这些值将与竞争性均衡一致，如果条件式（71）和式（72）得到满足。现在让我们看一下使得如下式成立的子集

$$\frac{EV_Y^a}{EV_Y^b} = \frac{\lambda^b}{\lambda^a}, \text{一个常数} \tag{73}$$

（当然，这等价于式（64））。式（64）、式（67）和式（71）定义了关于不同自然状态的概率分布的四个线性方程。所以约束条件

$$EDp^2\theta = 0 \tag{74}$$

为一个额外的线性约束，该式将只会在一个零测度集上被满足，除非 $Dp \cdot p\theta$ 为关于 Dp，D 以及 $V_Y^a\ (p\theta - (1+r))$ 和 $V_Y^b\ (p\theta - (1+r))$ 的线性函数，而这只有当 $p\theta$ 为常数或者 $D \equiv 0$ 时才成立。

$$p\theta \text{为常数} \tag{69}$$

要么

$$D \equiv 0 \tag{70}$$

我们现在看到其中的每一个都意味着

（a）$p\theta$ 为常数，意味着 $z_2=0$，而这又意味着要么

$$s^a = s^b \tag{71}$$

所有个体都具有相同的偏好（这种情况在前面我们已经涉及了）；要么

$$\delta^a = 0 \tag{72}$$

对于 $\delta^a=0$，

$$\frac{V^a_{Yp}}{V_Y} = \frac{V^b_{Yp}}{V_Y} \tag{73}$$

即既然不存在风险收入，那么

$$V^a_p = -C^a_2 V^a_Y$$

$$-\frac{pV^a_{pY}}{V_Y} = s(1-s^a)(1-\rho^a)$$

所以，我们要求

$$(1-s^a)(1-\rho^a) = (1-s^b)(1-\rho^b) \tag{74}$$

（b）$D\equiv0$。

为分析的简单，我们将分析的情况限制在常相对风险厌恶系数的范围内，从而有

$$\begin{aligned} V^j &= \phi^j(Y^j p^{s^j-1}) \\ &= \frac{[Y^j p^{s^j-1}]^{1-\rho^j}}{1-\rho^j} \\ &= \frac{[(Q^j+R^j p)p^{s^j-1}]^{1-\rho^j}}{1-s^j} \end{aligned} \tag{75}$$

（根据式（50））。

所以，如果 $D\equiv0$，需有

$$\begin{aligned} &\lambda^a p^{(s^a-1)(1-\rho^a)}(Q^a+R^a p)^{-\rho^a} \\ &= \lambda^b p^{(s^b-1)(1-\rho^b)}(Q^b+R^b p)^{-\rho^b} \end{aligned} \tag{76}$$

对所有的 p 成立，或者取对数后再微分，得到

$$-(1-s^a)(1-\rho^a) - \frac{\rho^a R^a p}{Q^a+R^a p}$$

$$= -(1-s^b)(1-\rho^b) - \frac{\rho^b R^b p}{Q^b + R^b \rho} \tag{77}$$

通乘以 $Q^b + R^b p$ 和 $Q^a + R^a p$，然后再关于 p 求二阶导，得到

$$-(1-s^a)(1-\rho^a)R^a R^b - \rho^a R^a R^b$$
$$= (1-s^b)(1-\rho^b)R^a R^b - \rho^b R^a R^b \tag{78}$$

或者

$$R^a R^b s^a(1-\rho^a) = R^a R^b s^b(1-\rho^b) \tag{79}$$

这意味着：

$$R^a \text{要么} R^b = 0$$

即

$$\delta^a\left[1 - \frac{\beta^a(s^a - s^b)}{\bar{s}}\right] = \frac{s_b}{\bar{s}}\delta^a \tag{80}$$

但是这要求 $\delta^a = 0$；或者

$$s^a(1-\rho^a) = s^b(1-\rho^b)$$

而经过一些转换后，仍然被证明需要有 $\delta^a = 0$ 的条件。

因而，在两种情况下均说明不存在市场风险。

简而言之，如果所有的人都具有单一的价格弹性和不同的偏好，除了在极特殊的情况下，关于利润不会有风险（p^* 为常数），市场均衡对于所有可能的技术形式都不会是一个受约束的帕累托最优。

在这一部分，我们得到了：当所有的个人都具有单一的价格弹性和不同的偏好时，如果不存在期货市场，那么市场将会是受约束最优且市场上不存在风险。而对期货市场的引进，增加了所有人收入的可变性（使用随机商品作为本位商品）；但是，目前我们还没有能够弄清期货市场的开放到底是一个帕累托改进，还是使所有个体的境况都变差的条件。

五、结论

本文可以被视做对起源于卡尔·波奇开创性工作的一系列研究的一个贡献。这类研究对于在不存在完全的风险市场条件下，市场机制对资源配置的有效性提出了怀疑。

近期的研究文献对以下几点提出了质疑：

（a）股票市值最大化标准的合理性；

（b）关于公司价值将与其规模成比例增加的假设，是否令人信服；

（c）关于市场机制是受约束帕累托最优的这个结果的普适性；

（d）现存市场集合的最优性。

本文考虑了这样一种情况，在该情况下技术使得公司股东期望公司的股市价值最大化，在该情况下，关于公司价值将随公司规模的增加而成比例增加的假设是合理的。然而，我们得到了一个很强的关于市场机制不是受约束帕累托最优的结论，而不是我们可以找到市场满足帕累托最优条件的“例子”之类的结论。相反，描述市场均衡特征的边际条件和描述最优的边际条件是不同的。我们已经证明了这两者重合的环境所要求的条件是相当苛刻的。一般而言，竞争性市场均衡既不具有交换有效性也不具有生产有效性。当所有的个体具有相同的偏好时，我们能够得到交换的有效性，但是，即使是这种限制都不足以保证投资配置的有效性。要达到这一点，我们需要所有的个体都是无差异的或者都具有单一价格弹性的条件。两种情况都说明股市在风险分担方面毫无作用。

在不完全市场条件下，价格有两个功能：它不但承担了传统的资源配置任务，它还在分担和转移风险（保险）方面扮演着重要角色。例如，当价格与数量反方向变化时，生产的风险就通过价格系统从生产者转移到了消费者。很显然，当我们对价格系统要求得太多，它的表现并不好，或者说，至少表现得没有有选择性的政府干预好。

（我们这里的结论，与诸如生产者不会承担充分的风险，因而市场就不能行之有效地对资源进行配置的幼稚观点是截然不同的。这一点是很重要的。）如果只存在不完全的风险市场，那么由于生产者而导致的风险（在存在完全的风险市场时，这是可以被保险的）会造成真实的社会成本；也就是说，在其他的制度安排环境下这种风险可能会被分散的事实，并不能改变如下事实：在我们所考虑的制度环境下，生产者确实承担着风险，而这种风险影响着他们的期望效用水平。关于这个问题的传统分析，将一种市场结构下的资源配置结果和另外一种市场结构下的可能出现的资源配置结果进行了错误的比较，关于这种比较我们在前面已经说明了它是不公平的，甚至是毫不相关的。在分析中，我们明确地考虑了在次优环境下，风险的真实成本。

我们这里所关注的无效性（以及 Stiglitz 1972a，1975b）为边际上的无效性，除此之外，注意到德里斯（Dreze，1974），哈特（Hart，1975）和斯蒂格利茨（1972a）所强调的，在那样的情况下有可能存在结构的无效性：给定其他企业的生产决定，每一家企业都做出了正确的决策；但是，还存在另外一个能导致帕累托改进的（相互一致的）行动集合，这

是非常重要的。换而言之，可能存在多重纳什均衡，其中的一些均衡帕累托占优另外的均衡。

关于我们的分析必须注意到两个明显的局限性。首先，我们没能够对扭曲的大小（或者它所导致的福利损失）做出定量估计；事实上，我们甚至都没有能够为分析偏离的方向提供一个清晰的定性分析框架（我们的分析确实对于哪些类的参数值可能是相关的提供了一些清楚的说明，比如，风险厌恶和偏好差别的大小）。虽然在原则上来说，我们可以设计出一组税收—补贴政策，以使经济能够达到受限帕累托最优，但是这种税收或者补贴的程度可能会随着产业的不同而显著变化，而且将取决于参数的大小（回报的概率分布），而对于这些任何程度上的准确估计都是很困难的。因此，这种税收—补贴政策体系的实际操作将会非常困难；因此从这种意义上说，我们的分析更多地应该被看作是，对市场能够提供对资源有效配置论点的正确性的质疑，而不应该被看作是为政府干预的描述性理论提供基础。

其次，关于正在运行中的市场集合的局限性，我们没有能够提供一个解释。我们这里所采用的分析方法：在给定的市场集合条件下分析经济的表现，然后评估是否存在最优的运行中的市场集合，显然是一种对于次优问题的次优分析方法：同时分析两个问题是可取的。而这正是我们研究计划的下一步。[10]

附录

关于收入的边际效用与状态无关的条件的推导：

条件 $dV_Y/d\theta=0$ 要求是非常苛刻的。为了说明该条件意味着什么，我们注意到

$$\frac{dV_Y^j}{dp} = V_{Yp}^j + V_{YY}^j(1-\alpha^j)\theta I^j \tag{A.1}$$

但是，根据罗尔恒等式有

$$V_p^j = -C_2^j V_Y^j$$

通过微分我们得到

$$V_{pY}^j = -C_2^j V_{YY} - V_Y^j \frac{dC_2^j}{dY} \tag{A.2}$$

10 在现在所进行的一些研究中，我们已经朝着这个方向做了一些努力，在不完全风险市场的原因为道德风险的情形下（与个人行为的不可观察性相联系），一般而言，市场均衡不是受约束的帕累托最优（Arnott and Stiglitz，1980）。

所以，根据式（A.1）和式（22）

$$\frac{dV_Y^j}{dp} = -V_Y^j\left[\frac{dC_2^j}{dY} + \left(\frac{-V_{YY}^j}{V_Y^j}\right)T^j\right] \tag{A.3}$$

令

$$\rho^j = \frac{-V_{YY}^j Y^j}{V_Y^j} = \text{相对风险厌恶系数} \tag{A.4}$$

$$\eta^j = \frac{d\ln C_2^j}{d\ln Y^j} = \text{需求弹性的衡量} \tag{A.5}$$

和

$$-\frac{1}{\sigma} \equiv \frac{\theta}{p}\frac{dp}{d\theta} = \frac{d\ln p}{d\ln X_2/X_1} \tag{A.6}$$

（σ为“加总的”替代弹性）。

然后（利用式（22）和式（A.3）至式（A.6））

$$\begin{aligned}\frac{dV_Y^j}{d\theta} &= \frac{dV_Y^j}{dp}\frac{dp}{d\theta} + V_{YY}^j(1-\alpha^j)pI^j \\ &= \frac{V_Y^j C_2^j p}{\theta Y^j}\left\{\left(\eta^j + \rho^j\frac{T^j}{C_2^j}\right)\frac{1}{\sigma} - \rho^j\left(1 + \frac{T^j}{C_2^j}\right)\right\} = 0\end{aligned} \tag{A.7}$$

当且仅当（设$\sigma \neq 1$），对于所有的θ有

$$\frac{T^j}{C_2^j} \equiv -\frac{\rho^j\sigma - \eta^j}{\rho^j(\sigma - 1)} \tag{A.8}$$

显然式（A.8）要求是非常苛刻的。例如，如果在收入弹性、相对风险厌恶系数和加总的替代弹性均为常数，那么要使式（A.8）对所有的θ均成立，必须有对所有的j，C_2^j与θ成比例，即商品2的总产出为每一个体所消费的比例，在任意自然状态下都是相同的。

符号系统

p：商品2的价格（相当于商品1）；

Y^j：个体j的收入（以商品1为本位商品）；

X_i：商品i的产出；

I_i：对产业i的投资；

I^j：个体j的初始财富；

I：总投资水平；

α^j：j个体的初始财富中投资在产业1的比例；

β^j：产业 2 为个体 j 所拥有的比例；

γ^j：产业 1 为个体 j 所拥有的比例；

C_i^j：个体 j 对商品 i 的需求；

C_i：对商品 i 的总需求；

θ：自然状态；

^：标于变量上方以表示在约束帕累托最优中的相应变量；

η：需求的收入弹性；

ρ：相对风险厌恶系数；

σ：总（需求）替代弹性。

参考文献

ARNOTT, R. and STIGLITZ, J. E. (1980), "Moral Hazard and the Inefficiency of Competitive Equilibrium" (mimeo, paper prepared for Conference at Bell Laboratories).

ARROW, K. J. (1964), "The Role of Securities in the Optimal Allocation of Risk-Bearing", *Review of Economic Studies*, 31, 91–96.

BORCH, K. (1962), "Equilibrium in a Reinsurance Market", *Econometrica*, 30, 424–444.

BORCH, K. (1960), "The Safety Loading of Reinsurance Premiums", *Skandinavisk Aktuarietidskrift*, 163–184.

BORCH, K. (1968), "General Equilibrium in the Economics of Uncertainty", in Borch, K. and Mossin, J. (eds.) *Risk and Uncertainty* (London: Macmillan) 47–58.

BORCH, K. (1968). *The Economics of Uncertainty* (Princeton University Press).

DIAMOND, P. (1967), "The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty", *American Economic Review*, 57, 759–776.

DRÈZE, J. H. (1974), "Investment under Private Ownership: Optimality, Equilibrium and Stability", in Drèze, J. H. (ed.) *Allocation under Uncertainty: Equilibrium and Optimality* (Macmillan) Chapter 9.

GROSSMAN, S. J. and STIGLITZ, J. E. (1977), "On Value Maximization and Alternative Objectives of the Firm", *Journal of Finance*, 32 (2), 389–402.

GROSSMAN, S. J. and STIGLITZ, J. E. (1980), "Stockholder Unanimity in Making Production and Financial Decisions", *Quarterly Journal of Economics*, 525–542.

HART, O. (1975), "On the Optimality of Equilibrium when the Market Structure is Incomplete", *Journal of Economic Theory*, 11, 418–443.

JENSEN, M. C. and LONG, J. B. Jr. (1972), "Corporate Investment under Uncertainty and Pareto Optimality in the Capital Markets", *Bell Journal of Economics and Management Science*, **3** (1), 151–174.

NEWBERY, D. and STIGLITZ, J. E. (1979), "Pareto Inferior Trade" (mimeo).

STIGLITZ, J. E. (1972a), "On the Optimality of the Stock Market Allocation of Investment", *Quarterly Journal of Economics*, 25–60.

STIGLITZ, J. E. (1972b), "Some Aspects of the Pure Theory of Corporate Finance Bankruptcies and Take-Overs", *Bell Journal of Economics and Management Science*, **2**, 458–482.

STIGLITZ, J. E. (1975a), "The Efficiency of Market Prices in Long Run Allocations in the Oil Industry", in Brannon, G. (ed.) *Studies in Energy Tax Policy* (Cambridge, Mass: Ballinger Publishing Co.) 55–99.

STIGLITZ, J. E. (1975b), "Reply to Mr. Singleton on 'Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Take-Overs'", *Bell Journal of Economics and Management Science*, **16** (2), 711–714.

帕累托次优贸易*

本文证明了，在两个竞争而具有风险的经济体（没有保险市场）之间，在帕累托意义上自由贸易比没有贸易更差。本文所用的模型非常简单，可以非常清楚地展示了在存在一系列不完全市场的情况下，价格在传递风险和分担风险方面所起的作用。同时，模型也可以极大地揭示由此引起的无效率。

自由贸易是帕累托最优的这一信念，是为数不多的经济原则之一，至少直到最近，这一信念都受到了近乎普遍地认同。本文旨于证明，这一信念可能不是那么有根据的。我们构建了一个简单的模型，该模型中不存在一系列完备的风险市场，但在其他任何方面，它能满足我们对竞争经济的传统假设。同时，该模型证明了，在帕累托意义上，自由贸易比没有贸易更差。

模型背后的基本思想非常简单。有两个国家（地区），它们都同时种植一种有风险性的农作物和一种无风险的农作物。两个地区的产出量是完全负相关的（我们可以很容易地把这个模型扩展到不相关甚至正相关的情况，只要不是完全相关的）。在没有贸易的情况下，只要产出量下降，价格就会上升。如果需求函数具有单位价格弹性，价格变化就为农民提供了完全的收入保险。在自由贸易的情况下，风险性农作物的产出量的变化会互相弥补，

---

* "Pareto Inferior Trade", *The Review of Economic Studies*, with David M. G. Newbery Vol. 51, No. 1 (Jan., 1984), 1－12。本文的研究部分受到美国国际开发署（United States Agency for International Development）的支持，它是 Newbery and Stiglitz (1977) 文章的附录 E 的拓展。我们需要感谢 L. Perez 以及他在美国国际开发署的同事给出的评论。同时，我们要感谢编者，感谢出席在沃里克（Warwick）、都柏林大学（University of Dublin）、世界银行（The World Bank）、普林斯顿（Princeton）举办的研讨会的参与者们，以及 Roy Ruffin 和 Robert Lindsey 的有益评论。本文得到了美国国家科学基金会、英国社会科学研究委员会（Social Science Research Council of the U. K.），以及 IBM 的资助。对这篇论文的修改是在 Newbery 离开剑桥大学在世界银行工作期间完成的。本文陈述的观点均为作者个人观点，并不代表世界银行或者美国国际开发署的观点。本文的较早版本包括对贸易政策的讨论，刊登在 Economic Theory Discussion Paper No. 23, July 1979, Cambridge, England, Econometric Research Program Research Memorandum No. 281, May 1981, Princeton, 以及 DRD Discussion Paper 50, January 1983, World Bank, Washington, DC。

使得价格稳定——价格不再为了抵消产出量的变化而变化。因此，种植风险性农作物获得的收益是变动的，从而农民面临的风险增加了。这导致农民减少了风险性农作物的种植，从而提高了风险性农作物的平均价格。因为消费者具有单位价格弹性，所以在两种农作物上花费的钱不变，因此当风险增加的时候，开放贸易时农民的平均收入是不变的。因此，农民福利必然是下降的，如图1所示。

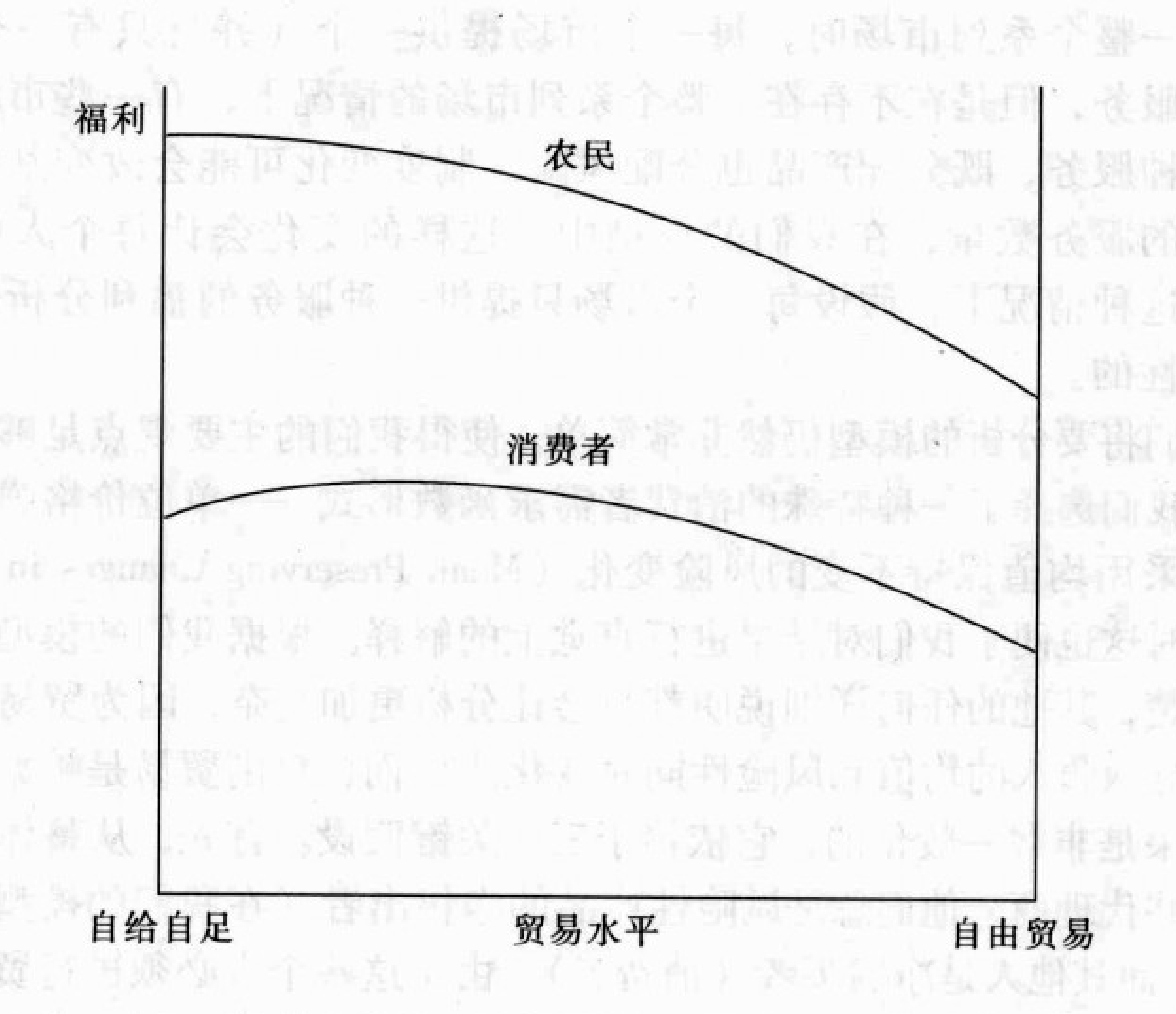


图1 开放贸易后的福利

然而，在开放贸易之前，消费者承担所有的风险。在自由贸易的时候，他们不承担任何风险。假如其他条件都相同，自由贸易将使他们福利改善。但是，风险性农作物风险的增加促使农民转为生产无风险农作物，使得风险性农作物的平均价格上升，从而导致消费者福利下降。在接近自给自足的情况下，风险收益主导了这种分配效应，如图1所示。但是在接近自由贸易的时候，情况正好相反。如果供给和价格的变化足够大（如果生产者是非常厌恶风险的，就能满足这一点），并且如果消费者的风险收益足够小（如果消费者不是那么厌恶风险，就能满足这一点），那么开放贸易就会使消费者福利状况恶化。既然生产者福利肯定会变差（在这个模型中），因此，自由贸易比起自给自足而言，是帕累托次优的。

我们的结果与标准的福利经济学定理（认为自由贸易是帕累托有效

的）的调和之处是非常明显的——传统观点不仅要求市场是竞争性的（同我们假设的一样），同时要求市场是完全的。在我们的模型中，必须有一系列保险市场来确保农民获得价格和产量的双重保险。由于多种原因——比如道德风险和逆向选择——这些市场是不完全存在的。在我们的例子中，自给自足为农民提供了收入保险，而开放一个新市场（对某一风险性商品进行国际贸易）与关闭这些暗自起作用的保险市场的效应是相同的。在存在一整个系列市场时，每一个市场提供一个（并且只有一个）可供交易的服务，但是在不存在一整个系列市场的情况下，有一些市场可能提供了多种服务，既分配产品也分配风险。制度变化可能会改变某个特定市场提供的服务数量，在我们的模型中，这样的变化会让每个人的福利变差。在这种情况下，假设每一个市场只提供一种服务的福利分析是具有严重误导性的。

我们将要分析的模型仍然非常简单，使得我们的主要观点足够清晰。尤其是，我们选择了一种特殊的消费者需求函数形式——单位价格弹性，以便于我们采用均值保持不变的风险变化（Mean Preserving Changes in Risk）概念，同时这也便于我们对结果进行直觉上的解释。根据我们的模型，一切都非常清楚，其他的任何详细说明都只会让分析更加复杂，因为贸易的变化一般都会导致收入的均值和风险性同时变化。[1] 然而，自由贸易是帕累托无效的这一结果是非常一般化的，它依赖于三个关键假设。首先，从整体看来必须存在一些代理商，他们总是风险性产品的净售出者（在我们的模型中，就是农民），而其他人是净购买者（消费者）。由于这些个人必须进行贸易，所以他们的福利受到价格分布的影响。[2] 第二，生产者和消费者都不能对其面临的风险（风险性农作物的产出和价格的变动）进行保险。贸易水平的变化改变了价格分布以及个人面临的风险。如果他们可以完全进行保险，那么风险就不会变化，从而就没有负面的结果。然而，我们认为更现实的假设是：个人无法为自己完全进行保险，因此风险和风险变化对结果有影响。第三，不同国家的产出量不是完全相关的，因此（根据我们的假设，不同国家在其他方面都是完全相同的）开放贸易不会影响价格分布。因此，尽管我们假设两个国家的产出量是完全负相关的，但只要不是完全正相关就可以。

1 在其他文章中（Newbery and Stiglitz，1982a），我们已经在一个更为一般的模型中证实了竞争（例如，自由贸易）均衡的无效率（除非在非常严格的条件下）。

2 因此，如果所有人的情况都是相同的，对我们的分析至关重要的价格分布的变化，就没有任何福利效果。

以上这三个假设，使我们的文章不同于以往关于贸易和不确定性的研究。大部分其他的研究都采用了代表性的消费者——生产者的概念，假设价格分布是由外生因素而不是内生因素决定的，同时假设不同国家的产出量是完全相关的，并且/或者重在研究存在完全风险市场的情况，在那样的情况下，农民可以分散风险。[3] 与我们的研究最为接近的是赫尔普曼和瑞兹（Helpman and Razin，1978a，b）的杰出成果，他们讨论了一般均衡效应。但是，他们假设存在一个完全的权益市场（一个比完全的风险市场稍弱的假设），并且假设存在完全相关，他们还采用了代表性的消费者——生产者的概念。[4]

本文分为两个部分。在第一部分，我们建立了基本的模型，并且推导出自由贸易是无贸易的次优情况。在第二部分，我们证明，该部分的结论比第一部分简单模型的结论更为有力。我们讨论了重要的假设，对模型进行了其他的解释，并且对本文分析的拓展方向给出了建议。

一、自给自足和自由贸易的比较

我们构建了一个简单的模型，在模型中有两个地区，它们在其他方面都是相同的，除了一点：它们的风险性农产品的产出是完全负相关的。每一个地区有 n 个完全相同的农民，m 个完全相同的消费者。我们首先来描述农民，然后是消费者，最后是有贸易和没有贸易两种情况下的市场均衡。然后，我们对两种均衡进行比较，并建立自由贸易是没有贸易的帕累托次优情况。

1. 生产者

n 个相同的生产者，每人拥有一单位土地。通常，一个农民将 x 比例的土地用来种植风险性农作物，用下标 r 表示，剩下的 $1-x$ 比例的土地用于种植无风险农作物，用下标 s 表示。风险性农作物的每英亩产出量为 θ，这

3 要了解更多的关于贸易和不确定性的近期文献，请参见 Pomery（1979）。那些集中关注价格分布外生的小国家情况的研究包括 Batra and Russell（1974，1975）以及 Ruffin（1974a，b）。这几篇文章都假设了生产者的产出量没有变化，而产出量的变化和价格的关系，在我们的分析中扮演了重要的角色。

4 即便存在完美的权益市场，消费者相同这一假设也是至关重要的。比如，Stiglitz（1982）证明，虽然在消费者相同的情况下，带有乘数不确定性和两个商品的股票市场均衡，总是受限的帕累托最优（constrained Pareto optimum）的，但是如果消费者不同（无论是他们对商品的偏好不同，还是他们对风险的偏好不同），就绝不是这样。

是一个随机变量，均值为1，方差为σ^2。无风险农作物的产出量总是1（这些都是标准化之后的）。

农民都是厌恶风险的；他们在知道θ（天气状况）之前决定x，最大化利润的期望效用：[5]

$$EU(\pi), U' > 0, U'' < 0 \tag{1}$$

其中，

$$\pi = xp\theta + q(1-x) \tag{2}$$

表示利润，p是风险性农作物的价格，q是无风险农作物的价格。

假设农民知道价格和θ之间的关系，但是由于每一个农民的力量都很小，他认为自己（在任何自然状态下）对价格没有影响。[6]因此，我们的分析集中于竞争的理性预期均衡。最大化期望效用的x由下式的解给出：

$$EU'(\pi)(p\theta - q) = 0 \tag{3}$$

2. 消费者

我们可以非常方便地用间接效用函数来表示代表性消费者情况：$V = V(I, p, q)$，其中I表示不变收入。[7]在大部分分析中，我们需要对该效用方程进行特别的参数化。我们令

$$V = \begin{cases} \dfrac{(Ip^{-a}q^{-b})^{1-\rho}}{1-\rho} & \text{当}\rho \neq 1 \quad (4) \\ \log I - a\log p - b\log q & \text{当}\rho = 1 \quad (5) \end{cases}$$

其中，ρ是相对风险厌恶系数。由此可以得到对两种农作物的总需求函数，它具有单位价格弹性和单位收入弹性：

$$Q_r = \frac{amI}{p}, Q_s = \frac{bmI}{q} \tag{6}$$

其中，Q_i是对农产品i（风险性的或是无风险的）的总需求，mI表示消费者的总收入。

我们对效用函数进行这样的特殊处理，有如下几个原因：它极大地简化了计算过程；这样的效用函数形式使得传统的福利经济学分析中关于消费者

5 本分析的一处无关紧要的简化是：生产者并不消费自己生产的产品，他们的福利仅仅取决于自己出售农作物的收入。

6 在我们的简单模型中，因为价格变化的唯一来源是供给变化，因此价格就是受θ决定的一个函数；在更一般化的模型中，它会是一个随机函数。

7 I是消费者拥有的第三样商品的禀赋（或者是一个希克斯复合商品（Hicksian composite commodity），代表所有其他的商品），作为记账单位使用。

剩余的计算是有效的；[8] 这意味着，我们所考虑的政策变化不会产生收入重新分配的效应，[9] 这样的特殊化使得我们可以相当容易地将贸易政策的效率和分配结果区分开来。另外，具有单位价格弹性的需求函数在分析多产品风险的过程中，起着重要的分界线作用；如果两种商品的替代弹性少于1，那么价格变化必然导致：农民认为无风险商品比风险性商品的风险更高（请参见 Stiglitz，1972）。最后，效用函数具有不变的相对（收入）风险厌恶系数，ρ，并且在式（5）表示的单位风险规避系数的情况下：

$$V_{Ip} = V_{Iq} = 0 \tag{7}$$

因此，价格变化并不影响收入的边际效用。

3. 没有贸易的市场均衡

该均衡的结构是非常简单的：

（a）给定生产者对价格和 θ 的关系的预期是正确的，生产者决定对农作物种植的土地分配 x。

（b）在收获农作物之后，农作物在市场上的交易是竞争性的。市场价格使得需求（由式（6）给出）等于供给（取决于 x 和 θ）。如果所有的农民采取相同的行动，则市场出清价格为：

$$p = \frac{ay}{x\theta}, q = \frac{by}{1-x} \tag{8}$$

其中，$y = mI/n$ 表示平摊到每个农民头上的消费者收入。式（8）意味着两种农作物都获得完全无风险的回报，而且，如果他们同时种植两种农作物的话，这两种回报也必然相同。

$$p\theta = q \tag{9}$$

式（8）和式（9）意味着

$$\frac{ay}{x} = \frac{by}{1-x} \tag{10}$$

并且由此得到了均衡解：

$$x = \frac{a}{a+b}, q = (a+b)y, p = (a+b)y/\theta \tag{11}$$

生产者的利润是 $\pi = q$（运用式（9）和式（2）），因此生产者的福利是 $U\{(a+b)y\}$。在式（5）的对数形式中，代表性消费者的平均福利是

8 请参见 Samuelson（1942）。

9 对于其他的效用函数，农民的平均收入会由于敞开贸易而增加或减少。参见本文的第二部分第2节。

$$EV = V_0 + aE\log\theta, V_0 = \log I - (a+b)\log\{(a+b)\gamma\} \quad (12)$$

其中 V_0 是没有风险时的效用。[10]

需要注意的是，单位弹性的消费者需求函数将所有的风险从农民身上转移到了消费者身上。对风险性农作物的波动性供给仅仅影响到其自身的价格，而不会影响到无风险农作物的价格。除这种特殊情况之外，一种商品的生产风险往往会扩散开来，会对其他产品造成价格风险（Stiglitz，1972）。

4. 自由贸易的市场均衡

现在假设，在山脉的东边，有另外一个地区，它在绝大部分方面都和上面描述的地区相同——除了天气状况不同之外。[11]当西边下雨的时候，东边是干旱的，反之亦然。两个地区的风险性农作物的产量是完全负相关的。我们正式令 $\theta^E + \theta^W = 2$。最初东西方之间没有贸易，但是山脉之间有一条通道，使得双方可以近乎无成本地进行贸易。如果贸易开放，就会有竞争性的自由贸易；如果没有贸易，两个地区就仍然维持自给自足的状况。自由贸易是否会带来好处？

很强的对称性假设确保总会出现一个对称性的均衡，但是可能还有另外一个非对称均衡：每一个国家都相对专业化地生产一种农作物。如果农民不是那么厌恶风险，就会只有一个唯一的对称均衡。本文的大部分分析都讨论此种情况。

在对称的自由贸易均衡中，每一个地区都将相同比例（x）的土地用于种植风险性农作物，$1-x$ 的土地用于种植无风险农作物，从而获得总供给：

$$Q_r = nx\theta + nx(2-\theta) = 2nx, Q_s = 2n(1-x) \quad (13)$$

价格会完全稳定在以下水平：

$$\bar{\bar{p}} = \frac{a\gamma}{x}, \bar{\bar{q}} = \frac{b\gamma}{1-x} \quad (14)$$

因此，现在的总利润是

10 θ 的随机性带来的损失（用美元表示）大约是（$a/2$）σ^2。在相对风险厌恶系数为常数，且不等于 1 的时候：

$$EV = E\frac{\theta^{a(1-\rho)}V_0}{1-\rho}, \ V_0 = \left\{\frac{n}{m\ (a+b)}\right\}^{(1-\rho)(a+b)} I^{(1-\rho)(1-a-b)}$$

风险带来的损失（用美元表示）大约为$\frac{1}{2}a\ \{1-a\ (1-\rho)\}\ \sigma^2$，这和 $\rho=1$ 时的对数情况相同。

11 这排除了传统的贸易原因，使得我们可以集中关注风险因素。在本模型中，地区只在天气状况上拥有比较优势。很明显，这强烈地证明了本研究得到的所有政策结论。参见第二部分第 4 节。

$$\pi = \bar{\bar{p}}x\theta + (1-x)\bar{\bar{q}} = (a\theta + b)y \tag{15}$$

并且，农民的收入是具有风险性的，尽管其均值和以前一样。在两种农作物之间的土地分配由式（3）的解给出：

$$EU'\{y(a\theta + b)\}\left(\frac{a\theta y}{x} - \frac{by}{1-x}\right) \equiv EM(\theta, x) = 0 \tag{16}$$

5. 自由贸易和自给自足的比较

将自由贸易情况下的均衡值表示为 $\{\bar{\bar{x}}, \bar{\bar{p}}, \bar{\bar{q}}\}$，自给自足情况下的均衡值表示为 $\{\hat{x}, \hat{p}, \hat{q}\}$。命题1比较了两种情况下的均衡配置：

命题1：如果农民是厌恶风险的，比起自给自足的情况，在自由贸易条件下无风险的农作物更便宜，用于种植风险性农作物的土地更少。

证明：在自由贸易以及农民厌恶风险的情况下，风险性农作物的平均回报一定超过无风险农作物的平均回报，从式（14）可以得到：

$$E\bar{\bar{p}}\theta = \bar{\bar{p}} > q \text{ 或} \frac{ay}{\bar{\bar{x}}} > \frac{by}{1-\bar{\bar{x}}}$$

与式（10）表示的自给自足均衡相比较，这意味着 $\hat{x} > \bar{\bar{x}}$，从而 $\bar{\bar{q}} < \hat{q}$。

更一般地，如果 $\Delta x = \hat{x} - \bar{\bar{x}}$ 表示从自给自足转换为自由贸易时所减少的分配给风险性农作物的土地份额。那么，命题2就给出了两个用于确定 Δx 大小的重要参数。

命题2：（a）农民越是厌恶风险，Δx 就越大（Diamond and Stiglitz, 1974）；（b）θ 的风险性越大（在 Rothschild - Stiglitz，1970 文中所定义的），Δx 就越大，只要

(i) θ 值的范围不是太大，或者

(ii) 相对风险厌恶系数 R 为常数，且 $R<1$。

任何一个条件都能保证 M（θ，x）是 θ 的凹函数，[12] 并且确保这个结果。

---

12 因为，$M \equiv U'[(a\theta + b)y]\left(\frac{ay}{x}\theta - \frac{by}{1-x}\right)$，并且 $R \equiv -\pi U''(\pi)/U'(\pi)$，

$$\frac{\partial M}{\partial \theta} = ayU'\left\{\frac{1}{x} - \frac{R}{a\theta + b}\left(\frac{a}{x}\theta - \frac{b}{1-x}\right)\right\}$$

$$\frac{\partial^2 M}{\partial \theta^2} = \frac{Ra^2U'y}{(a\theta+b)^2}\left[-\frac{a\theta}{x} - b\left(\frac{2}{x} + \frac{1}{1-x}\right) + \left\{R - \frac{R'\pi}{R}\right\}\left(\frac{a\theta}{x} - \frac{b}{1-x}\right)\right]$$

如果 θ 值的范围很小，$(a\theta/x) - b/(1-x)$ 也会很小，所以 $\partial^2 M/\partial\theta^2 < 0$。

或者，如果 $R' = 0$ 并且 $R \leqslant 1$，则 $\partial^2 M/\partial\theta^2 < 0$。很清楚，这些条件可以变弱。

6. 福利分析

以下结果可以直接由式（15）得到：

命题3：开放贸易往往导致农民福利下降。在新的均衡中，收入均值不变，但是风险性增加了。[13]

现在，消费者不再面临任何风险。这使得他们福利改善，但是反过来，由于 x 发生了变化，这又使得他们福利下降了。要证明这一点，我们需要知道怎样的土地分配比例 x 可以最大化消费者福利

$$V = V\left\{I\left(\frac{a\gamma}{x}\right)^{-a}\left(\frac{b\gamma}{1-x}\right)^{-b}Z(\theta)\right\} \tag{17}$$

其中 $Z(\theta)>0$ 取决于贸易政策。因此，在自给自足情况下 $Z=\theta^{a}$，而在自由贸易的情况下 $Z=1$。然而，对于任何满足 $Z(\theta)>0$ 的值，我们可以观察到：

$$x = \frac{a}{a+b} = \hat{x} \tag{18}$$

是农民在自给自足情况下的决策。任何其他的决策都会让消费者福利下降。因此，从自给自足转变为自由贸易所造成的消费者福利的变化，取决于资源变化（Δx）的大小，同时也取决于消除消费者面临的风险而为消费者带来的收益大小。

根据命题2，Δx 取决于风险的大小以及农民厌恶风险的程度，而风险减少给消费者带来的收益取决于风险的大小，以及消费者对风险厌恶的程度。

因此，如果农民是足够风险厌恶的，那么自由贸易似乎让消费者福利变差了，如果将对称的自由贸易均衡和自给自足的状况相比较，就会发现这确实是正确的。然而，在农民的风险厌恶的临界值之上，对称的自由贸易均衡不再是唯一的均衡了，此时出现了多重均衡。很典型地，多重均衡的存在引发了稳定性问题，尽管稳定性显然只能在特定的调整过程中才能检验。在其他文章里（Newbery and Stiglitz，1982c），我们证明了，对于一个合理的调整过程，如果对称均衡是唯一的均衡，那么它就是稳定的，如果不是唯一的均衡，那么它就是不稳定的。毫无疑问，任何特定的调整过程似乎都是特别针对某些读者的，因此，我们可以通过求证如下事实来回避这个问题：是否存在一个农民风险厌恶的临界值，在该值之下，对称均衡是唯一的均衡，并且在此临界值，自由贸易在帕累托意义上比自给自

13　对福利损失进行数量估计，以它们占平均收入的比例来看，可以用泰勒级数近似来表示：

$$\frac{\Delta U}{(a+b)\ yEU'} = \frac{U\{(a+b)\ \gamma\}-EU\{(a\theta+b)\ \gamma\}}{(a+b)\ yEU'} \approx \frac{1}{2}\left(\frac{a}{a+b}\right)^{2}R\sigma^{2}$$

足更差。

既然农民在自由贸易均衡下福利总是变差，如果消费者在自由贸易下福利也变差，则自由贸易在帕累托意义上比自给自足更差的。即，如果

$$E\frac{(\bar{\bar{p}}^{-a}\bar{\bar{q}}^{-b})^{1-\rho}}{1-\rho} < E\frac{(\hat{p}^{-a}\hat{q}^{-b})^{1-\rho}}{1-\rho} \tag{19}$$

在对称均衡中，将价格替换掉，这个式子等价于

$$\frac{\bar{\bar{x}}}{\hat{x}}\frac{1-\bar{\bar{x}}^{b/a}}{1-\hat{x}} < \{E\theta^{a(1-\rho)}\}^{1/a(1-\rho)} \tag{20}$$

式子的右边是消费者的相对风险厌恶系数 ρ 的单调递减函数，而式子的左边是农民的相对风险厌恶系数 R 的单调递减函数。因此，存在一个单调递增的临界值方程 $R=f(\rho)$，使得消费者对于自给自足均衡和对称的自由贸易均衡无差异。可以证明 $f(0) \equiv R_0>0$，其中 R_0 的值取决于效用函数的形式、风险的性质，以及其他的参数。关键的问题是：是否满足 $R_0<R^*$，R^* 是使得对称均衡不再是唯一的均衡的临界值。

R_0 和 R^* 的值取决于参数 a，b，θ 的分布，以及农民的效用函数的形式，但是对于绝对风险厌恶系数或相对风险厌恶系数不变的效用函数，以及 $\theta=1\pm\sigma$ 的两点分布，很容易计算出它们的值。[14]

对于 $\sigma=0.9$，a 和 b 的值在很大范围内能保证 $R_0<R^*$：$0.2\leqslant a\leqslant 0.8$；$a+b\leqslant 0.9$，它们对于绝对风险厌恶系数不变的效用函数和相对风险厌恶系数不变的效用函数都成立。R_0 随着 b 的增加而增加，随着 a 的增加而减少，并且对于 $a\geqslant 0.4$ 的情况，R_0 小于3.0。然而，对于 $\sigma\leqslant 0.8$，很少有这样的情况，因此，至于自由贸易在帕累托意义上比自给自足差，风险程度必须是很极端化的。

这些计算结果允许我们断定：

命题4：存在一系列参数值使得唯一的自由贸易均衡是自给自足经济的帕累托次优状况。

这一模拟说明，满足上面条件的（R，ρ）值的范围随着 σ、a 和 b 的增加而增加。还存在更大范围的（R，ρ）值，使得自由贸易是潜在次优于自给自足的——在这样的意义下：尽管消费者在自由贸易下福利改善，但是为弥补生产者由于接受自由贸易而造成的损失，他们必须支付一笔固定费用，但他们不愿意支付这笔钱。

---

14 详细讨论可以向 Newbery 索取。

二、模型的拓展和局限

本文的前一部分使用了一个非常简单的模型，得到了一个非常强的结论：自由贸易可能是没有贸易的帕累托次优情况。这些结果在多大程度上是依赖于我们所使用的特殊的参数化方法？又在多大程度上依赖于我们使用的特定的制度假设？我们认为，我们在分析中采用的重要假设是：不存在完美的风险市场，尽管这个假设在模型中不是合理的，但是在实证上却是比传统采用的另一极端假设——风险市场是完美的——更可信的（至少在我们这里所探讨的情况下）。[15]

要探究为什么这个假设如此重要，让我们简单地考虑一下在另一种假设下的均衡的性质。

1. 金融市场

首先，我们注意到，在消费者具有对数效用函数的重要例子里，在没有贸易的情况下，是没有期货市场的；那么对于 $V_{IP}=0$，收入的边际效用同样是独立于自然状态的，因此市场均衡要求：如果引入小型的期货市场，这在保险精算上是公平的。但是对于生产者而言，利润是不变的，并且他们也要求存在保险精算上公平的期货市场。因此唯一的均衡是：在期货市场上没有交易。如果两个地区的产出量是完全负相关的，那么在自由贸易的情况下，也没有期货市场，因为价格不会改变。

开放一个交易农场所有权市场的后果更为严重（我们可以忽略所有道德风险问题，虽然它对于帮助我们理解为什么这样的市场可能不存在是非常重要的）。如果农民可以互相购买对方农场的股份，那么农民的收入就不会改变；因此，他们对于展开贸易和自给自足就会感到无差异，但是毫无疑问，消费者的福利改善了。因此，这就支持了标准的结论——如果存在证券市场的完全集合，那么有自由贸易比没有贸易更受欢迎。

2. 需求结构

我们采用单位价格弹性假设的主要原因是：它使得我们免于将传递效应和风险效应混淆起来。如果需求弹性不等于1，那么价格分布的均值保持变化就会改变农民的平均收入（Newbery and Stiglitz，1981）。尤其是，如果需

15 巴西的咖啡种植者通常并不拥有肯尼亚咖啡农场的股份，反之亦然。尽管我们没有解释完美风险市场的缺失，但是不难构建一个与我们的分析相容的模型，同时该模型拥有这样的性质：存在非完美风险分担。

求曲线的弹性为常数且大于1，那么农民的收入就是消费量的凸函数。所以，对于任意的 x，自给自足情况下的平均收入比自由贸易情况下的平均收入高。尽管在自给自足的情况下，生产者的收入是变化的，但其变化幅度小于产出量的变化幅度，因此比自由贸易下收入的变化幅度小。因此，可以证明：(i) 从没有贸易的情况变化到稍微有一点贸易的情况，生产者的福利总是增加的；(ii) 假如弹性不是太大，对自由贸易状况稍微做一点限制，这将增加生产者的福利。

分析对消费者的效应更难一些。对消费者的效应有三个：(i) 风险效应：同以前一样，开放贸易降低了消费者的风险；(ii) 传递效应：由于价格分布的变化，消费者的平均花销增加了，这使得他们福利变差；(iii) 分配效应：由于传递效应，分配效应会更小，而且事实上它可能是正的而不是负的。

类似地，如果需求弹性小于1，价格变化的降低将减少生产者的平均收入，如果弹性大于0.5，敞开贸易增加了生产者收入的变动性。因此，毫无疑问，生产者的福利变差了。消费者的效应也更加复杂：现在，传递效应是正的，但是分配效应更大。因此，存在很多种可能的模式。

如果价格弹性非常小（小于0.5），那么贸易前收入的变动性比贸易后收入的变动性更大，在这种情况下，即便没有直接的补偿安排，开放贸易也是帕累托较优的。

然而，对于需求弹性接近1的情况，定性意义上的特性仍然是不受影响的：如果农民是足够风险厌恶的，那么在帕累托意义上，自由贸易可能比没有贸易更差；如果需求弹性大于1，将没有贸易的状况稍微放松为有少量贸易，这是帕累托改进的；如果需求弹性小于1，生产者的福利是严格地下降的；如果生产者是足够风险厌恶的，那么对自由贸易的稍微限制是帕累托改进的。

3. 不完全相关的风险

推动我们的模型的三大重要假设之一是：不同国家的生产量不是完全相关的，否则开放贸易就不会影响价格分布。为了简化，我们采取了另一个极端的假设：完全负相关，因为这使得我们可以只用一个参数 θ 来描述世界的状态。然而，对于不完全相关的风险，结果依然成立，并且在其他文章中(Newbery and Stiglitz，1982c)，我们分析了对称风险的情况：处于状态 θ 的国家1和处于状态 θ' 的国家2是不可区别的，反之亦然。这就保证了产出量的不完全相关，同时保留了对称性，因此在分析上更加易于处理。进行了这样的拓展，我们就有可能检验不那么极端的贸易政策，而不是必须在自给自

足和完全自由贸易之间进行选择，并且我们可能证明，这两种极端情况都不是帕累托有效的。

4. 比较优势理论

传统贸易理论专家在很多场合下，都对我们的分析表示了怀疑，因为我们的分析看似忽略了比较优势原则这一传统贸易理论的基石。需要注意的是，这种观点是不正确的。尽管在我们的假设中没有长期的比较优势，而且每一个国家都具有完全相同的禀赋和偏好，但是在每一时期各国都有显著的比较优势，而这仅仅取决于天气。

事实上，我们认为这正是我们的分析的一个吸引人的特色：有大量证据表明，贸易在很大程度上不能仅仅靠要素禀赋或者偏好来解释。相应地，采用没有长期比较优势的贸易模型就是很合适的。[16]

然而，我们可以很容易地，将传统比较优势理论的因素融合到我们的分析中来。尽管我们此处不再正式地探讨这种综合，但我们可以简要地说明如何完成这一点。为了简化，假设存在两种要素，比如说资本和劳动力。假设无风险农作物是劳动力密集型的。为了简化，我们假设两个国家具有相同的要素禀赋。很清楚，我们的假设中还是没有长期比较优势。然而，我们可以探究，开放贸易的结果如何。我们之前的分析几乎同样适用，除了一点：现在农民并不面临着一条直的生产可能性曲线，而是一条凹的生产可能性曲线。然而，开放贸易仍然导致无风险农产品（劳动力密集型）的生产的增加。现在假设两个国家的要素禀赋稍有不同。现在就有两个效应：在没有风险的情况下，开放贸易会导致要素价格相等，一个国家的劳动力价格下降，它在另一个国家的价格就会上升。然而，现在还有一个风险效应，它将倾向于将生产转移到无风险农作物上来（大体上，在对无风险农作物生产具有比较优势的国家里，这种转移更加厉害）。这个效应可能会超过前一个效应，因此，两个国家的劳动力价格仍然都上升了；甚至有可能要素价格差别更大了。

三、结论

自由贸易（可以在不同国家的价钱上套利）可以稳定价格。缓冲库存（在不同日期的价格上套利）也可以稳定价格，因此我们的分析集中于这样

16　当然，还有其他可能的模型，来强调规模经济和垄断竞争；再一次的，在这样的模型中，敞开贸易未必是帕累托改进的。比如，可以参见 Arnott and Stiglitz（1981）。

一个问题：将稳定商品价格的职能交给私人投机者，这是否合适？

在经济学家中，流行着一种观点：如果市场是竞争性的，并且代理人具有完全信息，那么政府干预将导致无效率。人们运用这个观点来反对贸易限制，同时用来反对建立国际商品稳定性计划。

我们的分析质疑了这种论证的前提。我们证明了，没有依据表明，自由贸易是帕累托最优的。但是，我们的分析并不是指向相反的方向。我们已经证明了，如何分解改变市场结构造成的风险效应和分配效应，并且这两种效应可能是互相冲突的。贸易政策的改变（比如开放贸易）会对生产者承担的风险造成重大影响，生产者对此的回应是，改变他们的资源配置。由此产生的生产量模式的改变反过来影响了消费者，他们的风险也随着贸易政策发生了变化。同时，由此导致了风险从消费者身上转移到生产者身上，这本身使得消费者利益上升，而生产者做出的配置反应，会使得其自身情况变差。回到类似的缓冲库存情况，很有可能在两个日期之间敞开贸易，虽然完全稳定了价格，但是相对于没有贸易的情况，却是帕累托次优的，因此限制投机者的套利行为是有益的（请参见 Newbery and Stiglitz ，1981，1982b）。

当然，这并不是说：缺乏市场的完全集合自然意味着政府干预是合理的，但是，这确实说明，基于竞争的完全市场范例而呼吁自由贸易的简化论证，通常是很缺乏说服力的。

参考文献

ARNOTT, R. and STIGLITZ, J. E. (1982), "Equilibrium in Competitive Insurance Markets: The Welfare Effects of the Moral Hazard" (mimeo).

BATRA, R. N. and RUSSELL, W. R. (1974), "Gains from Trade Under Uncertainty", *American Economic Review*, **64** (6), 1040–1048.

BATRA, R. N. (1975), "Production Uncertainty and the Heckscher–Ohlin Theorem", *Review of Economic Studies*, **42**,259–268.

DASGUPTA, P. and STIGLITZ, J. E. (1977), "Tariffs vs. Quotas as Revenue-Raising Devices under Uncertainty", *American Economic Review*, **67**,975–981.

DIAMOND, P. A. and STIGLITZ, J. E. (1974), "Increases in Risk and Risk Aversion", *Journal of Economic Theory*, **8**, 337–360.

FISHELSON, G. and FLATTERS, F. (1975), "The (Non-) Equivalence of Optimal Tariffs and Quotas under Uncertainty", *Journal of International Economics*, **5**,385–393.

HELPMAN, E. and RAZIN, A. (1978a) "Uncertainty and International Trade in the Presence of Stock Markets", *Review of Economic Studies*, **45**(2), 239–250.

HELPMAN, E. and RAZIN, A. (1978b) *A Theory of International Trade Under Uncertainty* (New York).

HELPMAN, E. and RAZIN, A. (1980), "Protection and Uncertainty", *American Economic Review*, **70**(4), 716–731.

NEWBERY, D. M. G. and STIGLITZ, J. E. (1977) *The Economic Impact of Commodity Price Stabilization* (Washington, D.C.: USAID).

NEWBERY, D. M. G. and STIGLITZ, J. E. (1981) *Theory of Commodity Price Stabilization* (Oxford University Press).

NEWBERY, D. M. G. and STIGLITZ, J. E. (1982a), "The Choice of Techniques and the Optimality of Market Equilibrium with Rational Expectations", *Journal of Political Economy*, **90**(2), 222–246.

NEWBERY, D. M. G. and STIGLITZ, J. E. (1982b), "Risk Aversion, Supply Response, and the Optimality of Random Prices: A Diagrammatic Analysis", *Quarterly Journal of Economics,* **97**(1), 1–26.
NEWBERY, D. M. G. and STIGLITZ, J. E. (1982c), "Pareto Inferior Trade and Optimal Trade Policy" (Econometric Research Program Memorandum No. 281, Princeton).
POMERY, J. (1979), "Uncertainty and International Trade", Chapter 4 of Dornbusch and Frankel (eds.) *International Economic Policy.*
ROTHSCHILD, M. and STIGLITZ, J. E. (1970), "Increasing Risk I: A Definition", *Journal of Economic Theory,* **2**,225–243.
RUFFIN, R. J. (1974a), "International Trade under Uncertainty", *Journal of International Economics,* **4**, 243–259.
RUFFIN, R. J. (1974b), "Comparative Advantage under Uncertainty", *Journal of International Economics,* **4**, 261–273.
SAMUELSON, P. A. (1942), "Constancy of the Marginal Utility of Income", reprinted in *Collected Scientific Papers,* Vol. 1,37–53 (M.I.T. Press).
STIGLITZ, J. E. (1972), "Taxation, Risk Taking and the Allocation of Investment in a Risky Economy", M. C. Jensen (ed.) *Studies in the Theory of Capital Markets* (New York.)
STIGLITZ, J. E. (1982), "The Inefficiency of the Stock Market Equilibrium", *Review of Economic Studies,* **49**(2),241–261.

资本市场自由化、全球化与国际货币基金组织[1]

全球化的一个最有争议的方面是资本市场自由化——不是主导外来直接投资规则的自由化，而是影响短期资本流动（进出一个国家的投机性热钱）规则的自由化。在20世纪80、90年代，国际货币基金组织和美国财政部（US Treasury）力图推进全球资本市场的自由化，遇到了无数的反对。这种反对不仅来自发达国家，而且来自那些不太迷恋自由无约束市场教义的经济学家。市场原教旨主义者的自由市场教义，当时被众多的国际经济机构所鼓吹。20世纪90年代和新千年开始几年的经济危机，可以说部分甚至大部分起因于资本市场自由化——这一事实加强了对自由市场教义所持的保留态度。本文以最近的一篇国际货币基金组织论文作为出发点，探讨了为什么国际货币基金组织，在支持资本市场自由化方面错得如此离谱，以及为什么资本市场自由化会如此经常地导致经济更加不稳定，而不是经济的更快增长。

一、引言

最近的一篇国际货币基金组织会议论文，由国际货币基金组织前任首席经济学家，肯·罗格夫（Ken Rogoff）和他的合作者（Prasad et al. 2003）撰写，评估了金融市场整合的结果。该文作者还在《金融时报》（*Financial Times*）撰文（Rogoff and Prasad，2003），总结了他们的一些发现。这两篇文章在很多方面都是引人注目的。作者在文中探讨问题所持的严肃态度，以及公开质疑传统结论的勇气都值得赞扬。他们得出结论："一旦考虑进其他因素，比如贸易流和政治稳定性，则在金融整合和经济增长之间，很难得出有说服力的联系"。他们还发现，"那些努力进行金融整合的国家……面临

---

1 "Capital Markets and Economic Fluctuations in Capitalist Economies", *European Economic Review*, 36, North-Holland, 1992, pp. 269 – 306. 作者曾任世界银行的首席经济学家（1997 ~ 2000年），在那期间，作者积极地参与有关资本市场自由化的争论。作者要感谢研究助理 Francesco Brindisi，以及福特基金会、麦克阿瑟基金会和 Mott 基金会的资助。

更多的不稳定”。值得注意的不是他们说了什么——经济学界很多人长期以来已经得出了这种观点[2]——值得注意的是谁在说。事实上，如果他们得出任何其他结论，那才是真的令人震惊。

过去和现在的政策制定者应该认真阅读这份报告——包括美国财政部那些人，他们过去曾经，并且现在继续在推动发展中国家轻率地进行资本市场自由化；在最近的贸易谈判中，他们企图将资本市场自由化的要求作为投资协定（Investment Agreement，即所谓的新加坡问题之一）的一部分，并且坚持资本市场自由化应该成为同智利和新加坡的双边贸易协定的一部分。[3]

这项姗姗来迟的研究本身是对国际货币基金组织的一种批评，但是正如谚语所说：“有胜于无”。然而，作者在《金融时报》上的文章认为，这份报告更多是“演化性的而不是革命性的”，就显得缺乏诚意。罗格夫和他的合作者，不能如此轻易地掩盖国际货币基金组织改变其纲领的企图——在1997年9月香港会议上，国际货币基金组织强迫发展中国家实行资本市场自由化（后者并不情愿），两年以后，国际货币基金组织的总裁仍然继续号召实行资本市场自由化——虽然那时全球金融危机已经生动地证明了资本市场自由化的危险性。[4]

也许，被隔离在国际货币基金组织的研究部门内，上述文章的作者无法完全感受国际货币基金组织对其他国家开放资本市场所施加的压力，也感受不到在香港会议上，东亚的财政部长向我表达的那种恐惧——如果酝酿中的危机一旦成为现实，他们就将采取一些必要措施（也就是进行资本管制），在这种情况下，国际货币基金组织会做些什么。只有马来西亚愿意抵制国际货币基金组织的压力——是它的总理，而不是财政部长进行抵制。它意识到，不进行资本管制的风险超过了进行资本管制的风险。

2　参见巴格瓦迪（Bhagwati，1998），弗曼和斯蒂格利茨（Furman and Stiglitz，1998），罗德里克和弗拉斯克（Rodrik and Velasco，2000），斯蒂格利茨（2000，2002），罗德里克（2001），以及这些研究中引用的大批论文。

3　对这些条约中这项条款的批评，参见巴格瓦迪，丹尼尔·塔伦洛（Daniel Tarullo），约瑟夫·斯蒂格利茨和南茜·伯德莎（Nancy Birdsall）在众议院金融服务委员会以及国内和国际货币政策、贸易和技术小组委员会上的证词，“Opening Trade in Financial Services – The Chile and Singapore Examples”，1 April 2003（可参见以下网址 http：//financialservices. house. gov/）。财政部部长约翰·泰勒（John Taylor）在这些听证会上的证词表明，他或者没有完全吸取这些教训，并且（或者）意识形态和利益仍然主导着美国财政部（参见 www. imf. org/external/np/speeches/1999/092899. htm）。

4　Michel Camdessus，国际货币基金组织总裁，对基金执行董事会的演讲，1999年9月28日（参见 www. imf. org/external/np/speeches/1999/092899. htm）。

当国际货币基金组织提议改变它的纲领的时候，我问了一个简单的问题：有什么证据表明，资本市场自由化对那些国家有好处？世界银行的研究已经表明，这种自由化与经济不稳定存在系统相关——这不是辩论术（就像罗格夫和他的同事们所暗示的那样），而是坚实的计量经济学证据。[5]对像我这样来自学术界的人来说，国际货币基金组织无法提供金融自由化有利于增长的证据，以及不能反驳金融自由化不利于增长的证据，都是令人烦恼的。看起来国际货币基金组织并不相信政策应该基于理论或证据，或者是它有一个完全不同的议程——也许是为了增加金融市场的利益——或者它信奉政策应该以意识形态（这种意识形态与利益相吻合）为基础，而不是经济科学。当然，现代经济学已经越来越转向政治经济学问题——政治过程和政府政策之间的关系——因此，在现有国际货币基金组织的治理结构下，这样的结果并不令人感到奇怪。[6]

二、暗含的假设和隐藏的议程

这份报告最引人注目的地方，也许不是它的结论——如果作者得出任何其他结论，那会是一种巨大的知识上的不诚实——而是它的出发点。

1. “理论”说了些什么？

在这项研究的开始，作者反复声明“理论”预测到资本市场自由化有利于经济增长，并且能降低消费波动性。作者想当然地认为，读者会理解“理论”的含义：具有完美信息、完美资本市场以及完全竞争的新古典模型。但是这个模型对于发达国家是一个拙劣的描述，对于发展中国家和国际资本市场则是一个更加拙劣的描述。罗格夫本人应该已经意识到了这种“理论”的限制：他发表过一篇论文（Obstfeld and Rogoff，2000），论文详细地描述了几种很难与“理论”一致的典型事实，包括贸易和资产组合中的本国偏好，投资对于国民储蓄的依赖，[7]国家间消费的低相关性，真实汇率冲击的高度可变性和高持续性，以及汇率和宏观经济变量之间的弱关系。其他人注意到另外一些“理论”失效：利率平价等式似乎不成立以及资本

5 参见 Demirgüç－Kunt and Detragiache（2001）以及 Honohan（2001）。

6 在 Stiglitz（1999，2003a）中，我简要地探讨了这些问题。也可参见韦德（Wade，2002）。并不让人感到奇怪的是，国际货币基金组织的许多人并不同意这种政策分析。尽管这种分析对于理解国家层面的政府（公共）失灵是合适的，对于国际层面的问题似乎不合常规。

7 所谓的费尔德斯坦—霍里奥卡之谜（参见 Feldstein and Horioka，1980）。有关费尔德斯坦—霍里奥卡之谜的研究综述，可参见 Coakley 等人（1998）。

流动的顺经济周期性（可参见 Lewis，1995；World Bank，2000）。

2. 不完美信息

“理论”——在过去四分之一世纪中，不完美资本市场的大部分理论发展，在浩瀚的参考文献中并没有被引用——解释了为什么资本市场自由化会导致不稳定，而不是促进增长。没有能够理解并接受这些理论发展——它们在许多方面，与作为国际货币基金组织许多政策基础的市场原教旨主义相矛盾——也同样解释了为什么国际货币基金组织会采取早期的立场，当时“理论”（或者更准确一点，意识形态和利益）为政策提供了清晰指导，因而国际货币基金组织看不到寻找证据的必要性。信息不完美导致了信贷和股权配给。鉴于信息不完美在现代金融里面扮演的中心角色，作为一个关注国际金融的机构，国际货币基金组织对信息不完美的忽视，就显得更加令人震惊。

3. 认知混乱

考虑到写上述论文时和之前发生的事件，论文坚持以完全就业、完全信息和完全理性的新古典经济假设作为分析起点，就更加引人注目。在巴西，由于担心选举的结果，资本流出巴西，这不仅迫使政府向国际货币基金组织寻求帮助，还使得政府不得不将利率提升到非常高的水平，导致经济急剧下滑和高失业。注意：巴西并没有做任何事，证明这种恐惧是合理的；而且选举后国家的经济管理进一步表明，这种恐惧是多么的毫无道理。另一方面，巴西为了阻止资金流出所设定的高利率，对产出和就业造成了负面影响。[8] 对资本市场自由化的两种标准批评是：资本市场自由化更有可能引发危机，二者存在着系统相关性；同时资本市场自由化，损害了政府对不利宏观经济冲击进行反应的能力，因为它使得政府无法降低利率。这些都无法在一个假设无失业问题的“理论”中探讨。

在东亚危机中，国际货币基金组织和美国财政部，大声抱怨缺乏透明度——即信息不完美。危机本身部分是由贷款人拒绝展期贷款（roll over loans）引起的，而不是因为他们感到风险增加而要求提高利率——这恰恰是不完美和非对称信息理论早先讨论的信贷配给问题的典型例证。国际货币基金组织同时还抱怨过量的财务杠杆，然而在作为这篇文章理论基础的新古

8　1998 年，政府名义利率达到 28.6%，真实利率为 25.4%。当然，私人部门借款人不得不支付更多。增长率下降到 1998 年的 0.1%，1999 年的 0.8%；失业率增长到 1998 年的 9% 和 1999 年的 10%。1998 年 11 月到 1999 年 4 月的平均真实利率为 33.7%。参见 IMF，*International Financial Statistics*；World Bank，*World Development Indicators*；国际货币基金组织的独立评估办公室（Independent Evaluation Office，2003）。

典模型中，金融结构根本毫不重要。当然，在过量借贷问题上面国际货币基金组织是正确的，但那是因为金融结构很重要：破产会带来真实成本。新古典模型，罗格夫和他的同事所重复引用的“理论”，对这些问题几乎没有提供任何见解。[9]

4. 理性之外

此外，最近行为宏观经济学和金融方面的研究（Akerlof，2002）强调了非理性的重要性。正如查尔斯·金德伯格（Charles Kindleberger，2000）指出的，回顾经济危机的漫长历史，非理性与市场不完美（包括对信息不对称的利用）自从资本主义起源的时候，就是波动的主要原因。[10]尽管这些波动并不完全符合国际货币基金组织的“理论”，但它们是真真实实的。即便是在20世纪90年代，联邦储备委员会的主席阿兰·格林斯潘（Alan Greenspan）仍然提请人们注意“非理性的繁荣”，这种非理性被后来的事件所证实。即便国际货币基金组织也不止一次提到了超调（overshooting）——发生危机后的一种非理性的悲观心理，这似乎是国际货币基金组织的对汇率的干预主要理由之一。

5. 外国直接投资还是资本市场自由化

上述国际货币基金组织论文还有另外一个根本缺陷。它讨论了以总资本流动衡量的金融市场一体化，但没有对资本流动的类型进行充分地区分。对资本市场自由化的大部分批评，不是针对外国直接投资（FDI），而是针对短期资金流动。很多人害怕的是后者——它们会带来不稳定，无助于增长。总资本流动（包括短期和长期流动）无法产生希望的结果，使人们更加谴责短期资金流动：很久以来人们就认为外国直接投资对增长有正效应。如果总资本流动（长期和短期）的作用可以忽略，那意味着短期资本流动有负效应。当然，这与“理论”相一致——不是国际货币基金组织论文背后的那种天真“理论”，而是现代金融理论。但是，如果对不同类型资本流动的效应（包括区分不同类型的外国直接投资，特别是区分绿地投资和私有化，

9 当然，国际货币基金组织和亚洲金融政策的其他批评者意识到，更高的杠杆比率与更高的破产可能性相关，并且只要存在破产成本，企业财务结构就很重要。新古典模型忽视了破产成本的事实，是对莫迪格利安尼—米勒（Modigliani - Miller）新古典分析的最早批评之一（Stiglitz，1969）。但是更根本的是，当信息不完美和不对称时，对于财务结构的决定传达了信息并影响激励。

10 在某些情形下，很难区分哪些行为可以由非理性来更好的解释，哪些可以由信息不完美来更好的解释（例如，羊群效应）。参见 Banerjee（1992）；Bikhchandani et al.（1992）；Howitt and McAfee（1992）。

以及区分自然资源投资和其他投资）进行更进一步研究，那么分析会大大丰富。在这篇文章后面部分，我会回到这一点。

6. 资本市场自由化成功的前提条件

最后，鉴于资本市场自由化的正反两方面经验，那些希望进行资本市场自由化的国家，想要知道以下问题的答案：在哪些情形下资本市场自由化会带来好处？上述国际货币基金组织论文对于这个关键问题只提供了很少的指导。它赞同当前政策圈中的流行观点："好的治理"扮演了关键角色。它既没有定义"好的治理"的精确含义是什么，也没有解决多重共线性和逆向因果的重要问题——那些有好的治理的国家还会有许多其他特性，使得他们对于投资者更有吸引力，更能吸收与短期资本流动相关的冲击和不稳定。上述论文也没有提供一个国家列表——对这些国家来说，如果不改变现有的治理，资本市场自由化会是一种错误。但是显然理论——不是新自由主义教义背后的简单新古典理论，而是我们在前面提到过的更现实的理论——没有表明：好的治理本身会消除我们提到过的问题。好的治理并不会消除信息不完美，也不会消除非理性繁荣或者悲观心理。斯堪的纳维亚国家通常被视为有良好治理的国家，但这并没有阻止它们在10多年以前发生的重大危机；同样，美国通常也被看作拥有良好治理（虽然最近的丑闻使得这种断言显得有些苍白无力），然而美国也产生了经济泡沫，并且最终泡沫破裂。[11]美国经济足够强健，能够承担这些事件（尽管最近对美国数据的重新解释，包括破产企业名单的增加，表明经济衰退是第二次世界大战以后最严重的）；[12]美国可以进行刺激性的赤字融资，帮助经济走出萧条；并且在这些宏观经济波动中，短期资本流动起的作用相对较小。相反，我们很快会看到，发展中国家要虚弱得多，随着资本市场自由化而来的短期资本流动，一方面在导致经济波动方面起了重要作用，另一方面大大降低了政府抵消来自于其他方面波动的能力。

三、反对资本市场自由化的情形

我已经详细讨论了最近国际货币基金组织的那篇论文，部分是因为它有助于说明，简单模型与意识形态和利益结合，如何主导了（至少在官方层

11　即便在那以前，美国还发生过一次小型金融危机。1989年储贷协会的崩溃是顶峰，给美国纳税人造成的损失在1 000亿~2 000亿美元。

12　在泡沫形成过程中不仅有大量的资源错配，在泡沫破裂以后的产出损失——经济的潜在增长和实际增长之间的差别——也是巨大的。参见Stiglitz（2003b）。

面）全球化的讨论。简单模型断言自由市场一定会提高福利。事实上，长期以来我们已经知道，如果市场不完美，如果信息有限，或者如果市场不完全，则竞争性的市场均衡，一般来说不是“受约束的帕累托最优”。次优理论告诉我们，在存在其他市场不完美时，消除一种不完美（“使资本市场自由化”）可能不会带来福利改善。在本文的这一部分，我将超越这些一般的讨论，而更加具体地证明资本市场自由化怎样“在理论上”导致消费和产出的更大波动，以及更低的经济增长——简而言之，为什么资本市场自由化对于发展中国家是有害的。[13]

1. *为什么资本市场自由化会导致更大的消费波动*

在强调消费波动性方面，上述国际货币基金组织论文的确作出了重要贡献。标准的效用理论认为个人希望平滑消费（smooth consumption），根据“理论”，运作良好的资本市场可以使个人做到这一点。

如果短期资本流动的作用是平滑消费，那么当经济萧条时资本流入一个国家，而当经济繁荣时资本流出（至少是相对于稳态的流量）一个国家。事实上，任何一个观察者都会观察到资本流动（特别是短期资本流动）是顺经济周期的，而不是反经济周期的（World Bank，1999）。经常被引用的统计数字描述了在20世纪80年代（失去的十年）早期，拉丁美洲大多数国家艰难偿债。[14]考虑到资本流动特别是短期资本流动的顺经济周期性，要弄清楚资本流动如何发挥所谓的平滑消费的作用是很难的。如果消费波动性下降，罗格夫和他的合作者就不得不解释，在资本流动的顺经济周期性下，这是如何发生的。

显然，资本流动特别是短期资本流动（至少是在某些关键情形下）的顺经济周期性表明，标准“理论”是失败的。但这种顺周期性，完全与关于银行家的经典格言相一致——他们只借钱给那些不需要钱的人——并且与金融理论的新发展（在过去四分之一世纪中的发展）相一致，它们强调信贷配给和信贷市场中的其他不完美。在附录中，我们提供了与这些现象一致的简单模型。

13　更加广泛的补充性的讨论，可以参见 Initiative for Policy Dialogue 将要出版的关于资本市场自由化的书。斯蒂格利茨（2000，2002）以及 Ocampo and Martin（2003）。

14　在拉丁美洲和加勒比国家，1982—1984年每年的GDP增长率为0。净资源转移占GDP的百分比在1982—1984年为-3.8%，在1985—1990年为-3.4%。1981—1990年总的固定资本形成下降了2.4%。参见拉丁美洲与加勒比经济委员会（Economic Commission for Latin America and the Caribbean，1996）。

（短期）资本流动不平滑消费还有另外一个毫不令人奇怪的原因："理论"预测那些更有能力承担风险的人——汇率和利率波动的风险——会承担风险。但是，事实上，发展中国家却不得不承担这些波动的冲击，很多波动与它们本国发生的事情毫无关系。在20世纪70年代末期和80年代早期，美联储将利率提高到了前所未闻的水平，加剧了拉丁美洲的债务危机；即便它没有引发危机，也会对拉丁美洲国家有不利影响。而对于拉丁美洲国家，即使它们可以不受约束地从资本市场上借到资本，利率提高也会降低它们的"生命周期收入"，从而降低它们的消费；而在资本市场不完美时，它们的消费则降得低。

当人们察觉到一个国家会经受（时间长度不确定的）不利冲击（比如，政治动乱）时，事情会变得更加糟糕。贷款人会立即削减贷款，导致消费立即下降。这样，不利冲击就被放大了，消费波动性增加了。

当然，关于消费波动性增加或减少的两种观点，都有某种程度的合理性：金融市场一体化可能会使一个国家减缓小的波动，但在面临大的不利冲击时，金融市场一体化会使波动增加。在一般的凹性假设下，小冲击时平滑消费的收益，在程度上要小于大冲击时波动增加的损失。

2. 政治经济学：纪律和其他

关于资本市场自由化有利于增长的一个标准论据是，它提供了"纪律"。这种观点的支持者显然对于民主缺乏信心；他们不相信投票者能够选择出合适的领导者，从而提高投票者的经济利益（指广泛定义的经济利益，包含在一个更广泛的社会议程中），他们认为最好依靠华尔街金融家的判断。这被称为"市场施加的纪律"。后面，我将表明这种观点的反复无常——它们的易变性——以及它们的极端短视事实上对于长期增长和稳定会有负面影响。实际上，如果对"缺乏纪律"的担忧是合理的，金融市场就会带来更多的不利影响。例如，在20世纪90年代的早期，在整个拉丁美洲，（由资本市场自由化和私有化推动的）资本流动为消费（包括公共和私人消费）的迅速增加进行融资。经济情况的其他指标（应该注意到，对国外的负债增加，以及资产所有权转移给外国人的事实）已经警告大家：事情不像GDP显示的那么美好。如果政府目光短浅，它们会更倾向于利用繁荣时期的金融市场自由化，进一步增加消费并放松预算约束——而很少考虑未来的后果。随之而来的泡沫破裂，会导致预算约束变紧，但这个问题，很可能由下一届政府来面对。

3. 为什么资本市场自由化会导致整个经济更大波动

前面两节概括地指出：在给定产出波动水平的情况下，为什么金融市场一体化不会降低消费波动。但是我们不应该将产出波动视为既定的：对资本市场自由化的主要批评是，它会增加产出波动。

任何熟悉最近几十年来东亚和拉丁美洲所发生事件的人都可以看到，资本市场自由化对经济不稳定的作用。“热钱”涌入一个国家，经常是为无节制的消费融资，然后又涌出这个国家；当资本离开的时候，金融机构被削弱，经常走向破产，本国货币急剧贬值，给美元债务的偿还带来沉重压力。在资本流入过程中，汇率升值，使得进口竞争部门和出口部门陷入困境。某些政府（20 世纪 90 年代中期的泰国）试图防止这种事情发生，同时试图避免经济过热。这使得政府不得不削减高回报的公共投资并提高利率，除进入房地产投机以外的投资因此减少。在资本流出期间，金融机构被摧毁，信贷缺乏导致经济衰退。人们原以为，上述国际货币基金组织论文应该从这些众所周知的现象开始分析，努力分析那些导致资本流出（包括由国内和国外事件引起的资本流出）的情况。

在下面的附录中，我们描述两个模型，其中资本市场自由化会导致整体经济的更大波动，以及消费的更大波动。第二个模型是一个完全的新古典模型，另外一点是，我们使用了重叠代模型（overlapping generation model）而不是代表一代理人模型。在模型中，我们看到资本市场自由化如何导致国家面临新的冲击，并降低了价格机制提供的经济内在的吸收冲击的能力。好的年份意味着工资会很高，这会导致下一年份更高的资本存量，资本增加会提高工资，但是会降低利率。后者抑制了消费增加，而前者将正向冲击的收益（以及负向冲击的成本）散布给许多代理人。

第一个模型代表了对标准新古典模型的更根本的偏离，包括了导致信贷和股权配给的信息不对称性。为简单起见，我们将这种数量配给置于一个标准的凯恩斯主义模型中，在该模型中生产受总需求限制。同样，资本市场自由化会放大国内冲击（那些会增加生产率的冲击），同时降低内在的利率调整，提供的自动稳定器作用（在投资前景不好时，利率降低，从而减少储蓄，当投资前景很好时则相反）。

资本市场自由化导致经济的更大波动，还有另外一个重要原因：它使得政府无法使用反周期的货币政策。在东亚危机中，马来西亚可以避免提高利率，而那些执行国际货币基金组织计划的国家，则不得不提高利率，这不仅加重了经济衰退，还导致更多银行破产，从而使得结构调整的任务更加困难

且代价沉重。

4. 为什么资本市场自由化不会导致更快的增长或者更高的投资[15]

尽管上述国际货币基金组织论文试图说明资本市场自由化加快经济增长的某些途径，它并没有尝试检验相反的假设，而仅仅是说明其他解释的相对重要性。作为这篇论文起点的观点显然不成立，因为论文甚至都没有讨论资本市场自由化对增长不利影响的途径。因为我在别的地方已经更加详细地列出这些不同渠道（Stiglitz，2003c，d），这里我仅仅作概略的讨论。

（a）如果资本市场自由化导致更大的产出或者消费[16]不稳定（像前一节所表明的），那么它会提高企业投资所要求的风险溢价，从而阻碍投资。只要投资依赖于现金流和资产负债表效应，经济萧条就可能对投资存在特别大的负效应。

（b）由于（上面指出的）阻止资本外流需要更高的利率变动，使这些问题变得更加复杂。

（c）产出和利率波动限制了债务融资的使用。债务融资存在很大的负效应，特别是在股权市场发育不足的发展中国家；[17]这些都会导致资源配置的低效——产出更低——并且增长更慢。

（d）短期资本是高度短视的。因此经常受到赞扬的市场纪律会强化国家采取更加短视的政策，这会对增长造成负面影响。例如，资本市场经常强调预算赤字，而不询问钱是怎样花掉的；当一个国家被迫削减高回报的投资以平衡预算时，长期经济增长会受损害。

（e）各国越来越感到，为谨慎起见需要保持一定量的外汇储备，其数值等于外币计值的短期债务。但是持有这种储备存在很高的成本。通常它们是以美国（或者其他硬通货）国库券的形式持有，年收益率1%。这会有较高的机会成本，因为这种资金投资于经济的其他部门会有更高的收益。如果发展中国家的某家企业从美国某银行获得1亿美元的短期借款，同时支付20%的利率，那么这个国家必须以储备形式保留相应数量的货币，收益只有1%。结果，该国会出现对美国的净转移：本国的福利几乎

15　甚至是在国际货币基金组织报告以前，世界银行已经得出了同样的结论（World Bank，1999）。

16　二者都是独立的与资本市场自由化相关：消费不稳定特别与非贸易消费品部门的投资存在关联。

17　由于信息不对称和信息不完美，即便是在发达国家，通过发行新股权融资的投资也相对较少。

必然下降！[18]

最后，上述国际货币基金组织论文在许多方面似乎没有意识到一个关键问题：短期资本流动如何转化为更真实的投资。企业不能够（或者不应该）用可能迅速撤离的资金为长期投资融资（当然，消费可以用这种方式融资，特别是耐用品的购买）。

5. 外国直接投资

上述国际货币基金组织论文强调了金融市场的一体化——总的资本流动——包括外国直接投资。尽管人们（在国际货币基金组织以外）对短期流动不会导致经济增长，但的确使得经济更加不稳定存在广泛共识，但更多的人是赞同外国直接投资的。因此，金融市场一体化（包括外国直接投资）对增长没有很强的正效应的结果令人感到惊讶。

部分原因可能是，外国直接投资统计量中包含各种投资，有的投资并不会导致经济增长（或者至少是可持续增长）。我认为，如果使用准确的福利指标，投资与经济增长的正相关关系会更不显著。[19]例如，外国直接投资包括私有化。如果部分私有化收益用于消费，那么该国财富（国家作为总体所拥有的财富）减少，国家就更穷了。在某些情形下，私有化的确导致效率提高——例如，当存在运营亏损的公共企业时——但在其他情形下，则不是这样。[20]在许多发展中国家，许多外国直接投资集中在石油或者其他自然资源部门，通常的GDP度量指标没有考虑随着资源的消耗国家变得更穷的事实。此外，一大批并且还在不断增加的文献（资源诅咒悖论，参见Sachs and Warner，2001），解释了为什么自然资源开发常常与更快的经济增长没有关联。自然资源部门一般与经济的其他部分很少有联系；同时，引起汇率升值的“荷兰病”使得出口更困难，进口竞争部门的能力更弱。此外，外国企业可能进行贿赂，以一定的折扣得到自然资源（或者在制造品的生产中，甚至使用贿赂以取得保护或者垄断地位）。

在伴随着国际银行并购国内银行的金融市场一体化的情况下，其他引发负效应的原因是：国际银行可能不是那么愿意或不能贷款给国内的中小规模

18 本国福利提高的唯一情形是：它有效配置资本的能力比美国银行差得多。

19 这对于其他形式的资本流动也是正确的。例如，正如前面分析所表明的，如果以外币计值的短期借款与危机发生的可能性具有较高的系统相关性，那么福利导向的国民收入账户应该对考虑借款可能导致的损失精确估计值。

20 对于私有化影响的研究综述，参见Megginson and Netter（2001）。关于转型经济的综述，可参见Djankov and Murrell（2002）。

企业。[21]尽管在原则上，有可能使贷款更加稳定（因为风险更加分散），实际一些重要的例子（比如近年来在玻利维亚）表明：对国内银行体系的冲击，或者导致风险承担意愿变化的其他国内事件，都会导致信贷供给市场的收缩，从而引发发展中国家整体经济的收缩。[22]

6. 证明干预的合理性

短期资本流动对经济的其他方面——与资本流动不直接相关的部分——有很大负效应的事实，意味着外部性的存在。当存在外部性时，表面看来总是有政府干预的必要。问题仅仅是：是否存在只会消除外部性的负效应，而又不影响其他的干预措施；如果有这样的干预措施的话，最佳的干预形式是什么。智利和马来西亚的经验表明，这样的干预措施是存在的。[23]考虑到本文的篇幅，这里我不能够详细说明。

这里，我仅仅希望指出，即使国际货币基金组织也认识到，外部性在这个方面的重要性——证据是作者对“传染”影响的关注，以及用“传染”证明援助的合理性。但是如果经济危机表明政府行动是合理的，那么探讨引起危机的潜在原因就是有意义的（就像为了增进公众健康，人们不应该仅仅是建立更大的医院）。引起危机的原因包括使经济不稳定的短期资本流动。因此，努力稳定短期资本流动就是有意义的。即便干预是不完美的（也就是说，它们是“渗漏的”），降低短期资本流动的波动程度，也会带来巨大的社会收益。

四、结论

经济学家特别是发展中国家的经济学家，很久以来就对资本市场自由化的优点表现出质疑，可以参见 Diaz - Alejandro（1985）经常被引用的论文。在经济学内部，我没有见过任何一篇对这种质疑的综述。但是早在上述国际货币基金组织论文出现以前，我就怀疑这些质疑代表了一种共识。即便是劳伦斯·萨默斯（Lawrence Summers）在他去美国财政部以前，也表达过疑虑（Summers，1997）。正如我所指出的，资本流动的顺周期性质和资本市场的结构——使得发展中国家承担了利率风险和汇率风险——无疑使资本市场自

21 类似的考虑，也导致对美国银行跨州经营的限制，这种限制在1995年才被取消。

22 外国贷款者对于发展中国家的风险更加缺乏信息的事实，可能意味着存在更大的可能，发展中国家的某些类型的贷款，会附有“红字条款”，即实行数量配给。

23 中国和印度是另外两个例子。印度的情形，可以参见 Joshi（2003），也可参见 Wyplosz（2002）。

由化会带来巨大的风险。要不是罗格夫等人的论文得出如下结论：平均来说资本市场自由化降低了消费波动（相对于产出波动），它仍然不会全面地回答对资本市场自由化的批评，他们的注意力不是集中于“平均”，而是集中于极端事件。但是一篇明显偏向资本市场自由化的文章——它的分析基于与许多（如果不是大多数）发展中国家情况不一致的理论假设——却得出这样一种怀疑自由化的结论，这对于资本市场自由化的争论是一个重大贡献。国际货币基金组织应该由强迫发展中国家实现资本市场自由化，转而与发展中国家一起设计更有效的资本市场干预措施，以便稳定资本流动，或者最好保证资本逆周期流动。国际货币基金组织应该更努力地对付资本市场的潜在失灵，设计出一些方法，使得更多的利率和汇率波动风险转移给发达国家和国际金融机构。并且，在未来的发展政策议程时，需要更多地依赖于证据，而更少地依赖于意识形态。国际货币基金组织在资本市场自由化方面的立场，已经破坏了它在许多团体中的信誉；尽管上述论文的作者声称论文代表国际货币基金组织思维的“一种演化，而不是革命”，但这篇论文仍然证实了，发展中国家中很多人长久以来就已经知道的：国际货币基金组织在这个领域的许多建议，使许多国家面对风险，而不是收益。

附录 I

一个“制度转换”的简单模型（其中，金融市场整合导致收入和消费的波动性增加）

假设一个经济体有两种自然状态，θ_1 和 θ_2，经济随机地（以马尔柯夫方式）从一种状态转向另一种状态，概率为 π。在两种状态中都有两个项目，一个是有风险的，另一个是安全的，并且贷款人不能监督借款人将从事哪一个项目。项目需要两期才能成熟，并为了简单起见，我们假设所有的贷款利率可变，第二期的利率等于当期流行的贷款利率。我们集中讨论均衡路径上的行为。因此贷款人和借款人对于下一时期的贷款利率会有（理性）预期，它们取决于状态。令 $\beta_i^j(r_i, r_{i'})$，$j=S$，R，代表一个借款为安全(S)，或者风险（R）项目融资的投资者的预期回报，此时经济状态为 i，利率为 r，并且预期到如果存在制度变化，则利率会是 r_{i}。（如果下一期状态相同，则合理地假设下一期的利率与这一期相同）在时期 1，借款人做出决定从事哪种项目，但是项目的结果由第 2 期的自然状态决定。当然，贷款人和借款人都知道今天和明天之间状态发生变化的可能性。

我们可以描述一个利率集合，在此集合内，在两种状态中都会从事安全

的项目：状态依赖的集合 $\{r_1, r_2\}_i$ 使得

$$\{r_1, r_2\}_1 : \beta_1^s(r_1, r_2) \geqslant \beta_1^r(r_1, r_2)$$

$$\{r_1, r_2\}_2 : \beta_2^s(r_2, r_1) \geqslant \beta_2^r(r_2, r_1)$$

我们令 ρ_i (r_1, r_2) 代表利率等于 $\{r_1, r_2\}$ 时，贷款人在状态 i 所作贷款的预期收益（贷款人会推断出借款人会从事哪种项目）。令 ρ^* = 无风险的国际回报率。我们假设状态 1 是坏的状态，并且

$$\text{Max}\ \rho_1(r_1, r_2) < \rho^* < \text{Max}\ \rho_2(r_1, r_2)$$

使得在状态 1，不管利率为多少，贷款人的预期回报非常小，使得在存在金融一体化的国家里不会有贷款发生。而在状态 2 会有贷款发生。分析的“技巧”是在一个封闭经济中，无风险的利率会随着经济的状态调整，使得某些贷款即便是在状态 1 也会发生。因此，产出的波动会更小。

我们假设有 M^* 个安全项目（每一个花费 1 美元），[24] 当借款人从事安全项目时，贷款人的预期回报达到最大。事实上，当借款人从事风险项目时，贷款人的预期回报如此之低，以至于他们不会进行贷款。该国的信贷文化非常薄弱，以至于国际贷款人不能够直接向消费者贷款（或者，他们不能够将资本家从骗子中间区分开来，并且有一个最大利率，在此利率之上资本家就不会借款，因此国际贷款人的预期回报为负。从而，不会有消费贷款）。

我们使用标准的凯恩斯总需求宏观经济框架，我们通过以下方式进行简化，假设所储蓄的收入比例为 s，是预期回报 ρ 的函数：$s(\rho)$，其中 $s' > 0$，并假设出口固定，记为 X，进口等于收入的一定比例 m，政府支出固定，记为 G，从而有

$$Y_i = X + G + I/m + s(\rho_i)$$

（1）封闭的金融市场

在封闭的金融市场情况下，ρ 调整以使储蓄等于投资。在“好的状态”下，所有的项目都进行，因此 $I = M^*$。

这意味着，在好的状态，利率可以产生的储蓄为所有的投资项目融资；[25] 而在坏的状态下，利率是使借款人不从事风险项目的最大利率。[26] 我们

24　所有的投资都需要非贸易品（国内生产的商品）。

25　我们调整参数使得在好的状态下，利率使得借款人愿意从事安全项目。

26　就是说，如果利率增加使得资金供给等于好项目的潜在需求，借款人事实上会从事风险项目。

可以很容易地从联立方程式解出（r_i，Y_i）：

$$Y_2 = [X + G + M^*]/[m + s(\rho_2(r_1^0, r_2^0))]$$

这是状态2中标准的收入—支出等式；

$$M^* = s(\rho_2(r_1^0, r_2^0))Y_2 = s(\rho_2(r_1^0, r_2^0))[X + G + M^*]/[m + s(\rho_2(r_1^0, r_2^0))]$$

保证了在状态2储蓄等于投资；

$$Y_1 = [X + G + I_1]/[m + s(\rho_1(r_1^0, r_2^0))]$$

状态1中标准的收入—支出等式；

$$\{r_1^0, r_2^0\}: \beta_1^s(r_1^0, r_2^0) = \beta_1^r(r_1^0, r_2^0)$$

保证了安全项目在状态1中进行；并且其中

$$I_1 = s(\rho_1(r_1^0, r_2^0))Y_1 = s(\rho_1(r_1^0, r_2^0))[X + G + I_1]/[m + s(\rho_1(r_1^0, r_2^0))]$$
$$= s_1^*[X + G]/m$$

其中 r_i^0 代表状态 i 时的均衡利率。[27]很容易证明对于 $i = 1$，2

$$Y_i = [X + G]/m$$

没有产出波动。存在一些消费波动，因为储蓄在好的状态下必定更高（从而在好状态下消费事实上会更低）。两种状态下消费的差别等于投资的差别。注意在此模型更一般的版本中，除了消费波动，也会有产出的波动。例如，如果 m，X 或者 G 对于自然状态或者真实利率敏感的话，会有产出波动。

（2）开放的金融市场

在完全开放的金融市场下，投资和储蓄的分离几乎是完全的——除了投资对收入的影响。存在一个国际利率决定储蓄；我们用 s^{**} 代表这个储蓄率。在好的状态下，所有的投资项目都被进行；在坏的状态下，不存在任何利率使投资盈利。因此

$$Y_2 = X + G + M^*/m + s^{**}$$

并且

$$Y_1 = X + G/m + s^{**}$$

开放资本市场会降低坏状态下的收入，并且（假设在好的状态，国家存在“资本稀缺”，即从国外借款）增加好状态下的收入（因为稀缺会导致更高储蓄和更低消费）。资本市场自由化因此会增加产出的波动。总体的消费波动增加还是减少则是不清楚的；两个状态中的消费差别现在是（1 -

27 在信贷配给的均衡中，r_i^0 使得 β_1^s（r_1^0，r_2^0）$=\beta_1^r$（r_1^0，r_2^0）。

$s^{**})(Y_2 - Y_1) = [(1-s^{**})/(m+s^{**})]M^*$，而在封闭经济模型中，两种状态下的消费差别是 $I_1 - M^*$。

这个模型也和观察到的资本流动的模式相一致——资本在好的时期流入国内，在坏的时期，流出国外（所有的储蓄都在国外投资）。

附录Ⅱ

一个包含不完全风险市场的简单模型（金融市场的整合导致收入和消费波动增加，并且福利和增长下降）

纽伯利和斯蒂格利茨（Newbery and Stiglitz，1984）证明在缺乏风险市场时，贸易开放如何会导致帕累托劣的均衡。想法很简单：存在单位的需求弹性时，价格和产出反方向变化，从而价格体系提供了完美的收入保险（当需求弹性接近于1，价格体制仍会提供很好的收入保险）。然而，贸易开放减弱了价格和数量之间的这种负向联系，从而使生产者面临更大的风险。这会阻碍在高回报活动中的投资：探讨清楚一般均衡模型中的效应，有可能发现不仅生产者状态更差（由于面临更多的风险），而且消费者状态也更差。

可以用类似的模型证明在没有好的保险市场时，资本市场自由化如何导致福利下降（以及消费和收入波动增加）。假设一个国家面对的国际贷款率，r 是可变的。然而外国人只给企业贷款（因为他们不能区分好的和坏的家庭借款人）。个人生活两期，在第一期工作。预算约束为

$$C_1^t = w_t - s_t$$

$$C_2^t = [1 + r_{t+1}]s_t$$

其中上标 t 代表第 t 代人。

为简单起见，我们假设有对数效用函数

$$U^t = \ln C_1^t + \ln C_2^t$$

这意味着 $s = 0.5$，并且

$$EU^t = -2\ln 0.5 + 2\ln w_t + E\ln(1 + r_{t+1})$$

在下面的讨论中，我们集中讨论两种情形，一种情形下随机性的唯一来源是利率，另一种情形下是国内生产率；在每一种情形下，我们将封闭的和开放的资本市场作比较。我们先讨论唯一的不确定性来源是外部利率的情形。

为简单起见，我们假设，短期资本实际上被转化为投资品。$F(K_t, L_t)$ 是常数规模报酬的生产函数，$f(k_t)$ 是每个工人的产出，其中 $k_t = K_t/L_t$。我们假设所有的资本在每一期都完全折旧。在封闭经济中，

$$k_t = 0.5w_{t-1} = 0.5[f(k_{t-1}) - k_{t-1}f'(k_{t-1})] = 0.5g(k_{t-1})$$

其中 $g \equiv f - kf'$。

稳态定义为

$$k^* = 0.5g(k^*)$$

我们假设经济处于稳态。不存在工资、利率或者效用的波动。

我们现在考虑经济开放时，会有什么发生。现在，

$$k_t = f'^{-1}(1 + r_t) \equiv h(r_t)$$

其中 r_t 是随机变量。因此 $w_t = g(h(r_t))$ 是随机变量，且时期 t 的个人的预期效用为[28]

$$E_t U^t = -2\ln 0.5 + 2\ln w_t + E_t \ln(1 + r_{t+1})$$

并且事前的预期效用（即任何一代人的平均预期效用）为

$$E\{E_t U^t\} = -2\ln 0.5 + 2E\ln w_t + E[E_t \ln(1 + r_{t+1})]$$

由此可得，资本市场开放无疑会使消费波动性增加。

对社会福利的影响更加困难一点，因为这部分取决于国际市场平均利率和封闭市场利率间的关系，以及社会福利函数的性质。例如，如果我们用 r_c 代表封闭经济中的利率，并且假设国际市场上（$1 + r_t$）的对数平均数等于1加上国内封闭经济的利率的对数值，即 $E[\ln(1 + r_t)] = \ln(1 + r_c)$，并且还假设有功利主义的社会福利函数，则随着生产函数中劳动和资本之间的替代弹性小于或者大于1，社会福利会随之下降或者增加。[29] 如果社会福利函数是凹的（即厌恶风险或者厌恶不平等），则即便替代弹性大于1，只要不是太大，社会福利也会下降。如果开放经济中的平均利率等于 r_c，则社会福利更有可能下降。[30]

生产率的可变性

即便国内生产有波动，类似的结果也成立，例如，令国内生产函数形式为 $Q_t = \theta_t f(k_t)$。在一个封闭经济中，现在有了消费和产出的波动。对于对数型的效用函数，有

$$s_t = k_{t+1} = w_t/2 = \theta_t g(k_t)/2$$

并且

$$E_t U^t = -2\ln 0.5 + 2\ln w_t + E_t \ln(1 + r_{t+1})$$

---

28 在下面，$E_t(X)$ 代表给定 t 时可得到的信息，随机变量 X 的条件预期。

29 随着 $\ln w$ 是 $\ln(1 + r)$ 的凹或者凸函数，社会福利会随之下降或者增加。这很容易证明 $d\ln w / d\ln(1 + r) = -\alpha/(1 - \alpha)$，其中 α 是国民收入中资本所占的份额。马上可以得到结果。

30 $\ln(1 + r)$ 是 r 的凹函数。

$$= -2\ln 0.5 + 2\ln w_t + E_t \ln\theta_{t+1} + \ln f'(w_t/2)$$

并且[31]

$$E\{E_t U^t\} = -2\ln 0.5 + 3E[\ln\theta_t] + 2E[\ln g(k_t)] + E\ln f'[\theta_t g(k_t)/2]$$

在开放的资本市场情形，

$$k_t^* = f'^{-1}((1 + r_t)/\theta_t) = h(r_t/\theta_t)$$

并且

$$w_t = \theta_t[f(k_t^*) - k_t^* f'(k_t^*)] = \theta_t g(k_t^*) = \theta_t g(h(r_t/\theta_t))$$

从而有

$$E_t U^t = -2\ln 0.5 + 2\ln w_t + E_t \ln(1 + r_{t+1})$$

如果我们假设国际利率没有波动，那么

$$E\{E_t U^t\} = -2\ln 0.5 + 2E[\ln\theta_t] + 2E[\ln g(h(r/\theta_t))] + \ln(1 + r)$$

显然，开放资本市场带来工资的更大波动（当θ高的时候，国家可以更多借款，使工资提高；相反，当θ低的时候，资本流出这个国家）。与封闭经济不同，今天高水平的工资对于下一期的利率没有负面影响。因此，平均来说，消费和预期效用会比封闭经济中波动更大，并且如果社会福利函数对不平等/风险有足够大的厌恶，社会福利会下降。唯一的不同是当经济中生产率更高时，可以得到更多的资源。我们必须将这种收益与更大波动性导致的损失进行权衡。对此的全面分析超出了这个附录的范围。这里，我们只关注一种特殊的情形：有单位弹性的生产函数、功利主义的社会福利函数，以及一种正规化使得封闭经济中利率对数的期望等于开放经济中利率的对数，即

$$\ln\alpha - (1 - \alpha)E\ln k_c + E\ln\theta_t = \ln(1 + r_o)$$

因此，以通常的符号

$$E\ln w_c = \ln(1 - \alpha) + \alpha\ln k_c + E\ln\theta_t$$

同时在开放经济中，对于每一个θ

$$\ln\alpha - (1 - \alpha)\ln k_o + \ln\theta_t = \ln(1 + r_o)$$

从而有

$$\ln\alpha - (1 - \alpha)E\ln k_o + E\ln\theta_t = \ln(1 + r_o)$$

接着马上有

$$E\ln w_c = E\ln w_o$$

因此，预期效用相同。但是，如果个人效用为

31　假设 $\ln(\theta_i)$ 是鞅。

$$\ln(w_t/2) + \ln(w_t(1 + r_{t+1})/2)$$

并且如果封闭经济中 w 和 r 负相关，则在封闭经济中跨代间的效用会有更少的波动；因此，如果社会福利函数是凹的，则在资本市场自由化以后，社会福利下降了。

更一般的，随着资本市场自由化，θ 的变化会产生工资的巨大变化，并因此会导致生命期效用巨大变化，特别是在如果劳动所占份额很小的时候。[32,33]

很显然，资本市场自由化导致产出的方差增大。

这些结论背后的直觉是简单的。在资本市场开放以前，效用的变化小于工资变化，因为当有正的生产率冲击时，工资会很高；当工资高的时候，储蓄会高，从而利率变低；年轻一代的生命期消费不会随着生产率（产出和工资）提高而增加。并且，随着储蓄增加导致下一代的工资增加，生产率冲击的好处，将与未来各代相分享，这又转而使未来各代人受益。

另一方面，一旦资本市场自由化，第一种效应不再存在（因为利率不取决于国家自身的储蓄）。并且，更高 θ 的影响只能被当时工作的一代感受到（因此只影响两期的消费，即工作的那一代人生命中的两期）。

参考文献

Akerlof, G. (2002), 'Behavioral Macroeconomics and Macroeconomic Behavior', *American Economic Review*, 92(3), 411–33.

Banerjee, A. V. (1992), 'A Simple Model of Herd Behavior', *Quarterly Journal of Economics*, 107 (3), 797–817.

Bhagwati, J. (1998), 'The Capital Myth: The Difference Between Trade in Widgets and Dollars', *Foreign Affairs*, 77(3), 7–12.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992), 'A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades', *Journal of Political Economy*, 100(5), 992–1026.

Coakley, J., Kulasi, F., and Smith, R. (1998), 'The Feldstein–Horioka Puzzle and Capital Mobility: A Review', *International Journal of Finance and Economics*, 3(2), 169–88.

Demirgüç-Kunt, A., and Detragiache, E. (2001), 'Financial Liberalization and Financial Fragility', in G. Caprio, P. Honohan, and J. E. Stiglitz (eds), *Financial Liberalization: How Far, How Fast?*, Cambridge, Cambridge University Press, 96–122.

Diaz-Alejandro, C. F. (1985), 'Goodbye Financial Repression, Hello Financial Crash', *Jour-*

---

32 在开放经济中，$d\ln k/d\ln\theta = 1/\beta$，其中 $\beta = -(1/d\ln f'/d\ln k)$，因此 $d\ln w/d\ln\theta = 1 + \beta(\alpha/1-\alpha)(d\ln k/d\ln\theta) = 1/1-\alpha$。

33 封闭经济中工资和效用的变化取决于 θ 的随机过程。例如，假设 θ 可以取两个值，θ_1 和 θ_2。则稳态由方程式 $k_1^* = \theta_1 g(k_1^*)$ 和 $k_2^* = \theta_2 g(k_2^*)$ 定义。

nal of Development Economics, 19(1–2), 1–25.

Djankov, S., and Murrell, P. (2002), 'Enterprise Restructuring: A Quantitative Survey', *Journal of Economic Literature*, 40(3), 739–92.

Economic Commission for Latin America and the Caribbean (1996), *The Economic Experience of the Last 15 Years. Latin America and the Caribbean, 1980–1995*, Santiago de Chile, United Nations/Economic Commission for Latin America and the Caribbean.

Feldstein, M., and Horioka, C. (1980), 'Domestic Savings and International Capital Flows', *The Economic Journal*, 90(358), 314–29.

Furman, J., and Stiglitz, J. E. (1998), 'Economic Crises: Evidence and Insights from East Asia', *Brookings Papers on Economic Activity*, 2, 1–135.

Honohan, P. (2001), 'How Interest Rates Changed under Liberalization: A Statistical Review', in G. Caprio, P. Honohan, and J. E. Stiglitz (eds), *Financial Liberalization: How Far, How Fast?*, Cambridge, Cambridge University Press, 63–95.

Howitt, P., and McAfee, P. R. (1992), 'Animal Spirits', *American Economic Review*, 82(3), June, 493–507.

Independent Evaluation Office of the IMF (2003), *The IMF and Recent Capital Account Crises—Indonesia, Korea, Brazil*, Washington, DC, International Monetary Fund.

Joshi, V. (2003), 'India and the Impossible Trinity', *The World Economy*, 26(4), April, 555–83.

Kindleberger, C. P. (2000), *Manias, Panics and Crashes: A History of Financial Crises*, 4th edn, New York, John Wiley & Sons.

Lewis, K. K. (1995), 'Puzzles in International Financial Markets', in G. Grossman and K. Rogoff (eds), *Handbook of International Economics*, Vol. 3, Elsevier, 1913–71.

Megginson, W. L., and Netter, J. M. (2001), 'From State to Market: A Survey of Empirical Studies on Privatization', *Journal of Economic Literature*, 39(2), June, 321–89.

Newbery, D., and Stiglitz, J. E. (1984), 'Pareto Inferior Trade', *Review of Economic Studies*, 51(1), January, 1–12.

Obstfeld, M., and Rogoff, K. (2000), 'The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?', in B. S. Bernanke and K. Rogoff (eds), *NBER Macroeconomics Annual 2000*, Cambridge, MA, MIT Press, 339–90.

Ocampo, J. O., and Martin, J. (eds) (2003), *Globalization and Development: A Latin American and Caribbean Perspective*, Stanford, CA, Stanford University Press.

Prasad, E., Rogoff, K., Wei, S.-J., and Kose, A. M. (2003), 'Effects of Financial Globalization on Developing Countries: Some Empirical Evidence', IMF Occasional Paper No. 220, September.

Rodrik, D. (2001), 'Development Strategies for the Next Century', in B. Pleskovic and N. Stern (eds), *Annual World Bank Conference on Development Economics 2000*, Washington, DC, World Bank.

— Velasco, A. (2000), 'Short-term Capital Flows', in B. Pleskovic and J. E. Stiglitz (eds), *Annual World Bank Conference on Development Economics 1999*, Washington, DC, World Bank.

Rogoff, K., and Prasad, E. (2003), 'The Emerging Truth of Going Global', *Financial Times*, 2 September, 21.

Sachs, J., and Warner, A. (2001), 'The Curse of Natural Resources', *European Economic Review*, 45(4–6), May, 827–38.

Stiglitz, J. E. (1969), 'A Re-examination of the Modigliani–Miller Theorem', *American Economic Review*, 59(5), 784–93.

— (1999), 'Knowledge for Development: Economic Science, Economic Policy, and Econo-

mic Advice', in B. Pleskovic and J. E. Stiglitz (eds), *Annual World Bank Conference on Development Economics 1998*, Washington, DC, World Bank, 9–58.

— (2000), 'Capital Market Liberalization, Economic Growth and Instability', *World Development*, 28(6), 1075–86.

— (2002), 'Capital Market Liberalization and Exchange Rate Regimes: Risk without Reward', *The Annals of the American Academy of Political and Social Science*, 579, January, 219–48.

— (2003*a*), 'Democratizing IMF and World Bank: Governance and Accountability', *Governance*, 16(1), January, 111–39.

— (2003*b*), *The Roaring Nineties*, New York, W. W. Norton.

— (2003*c*), 'Globalization and Growth in Emerging Markets and the New Economy', *Journal of Policy Modeling*, 25(5), 505–24.

— (2003*d*), 'Alternative Approaches to Stabilization Policy: An Overview', forthcoming in the book on capital market liberalization by the Initiative for Policy Dialogue.

Summers, L. (1997), 'The Case for Mild Repression', *The Economist*, 20 September.

Wade, R. H. (2002), 'US Hegemony and the World Bank: The Fight over People and Ideas', *Review of International Political Economy*, 9(2), summer, 215–43.

World Bank (1999), *Global Economic Prospects 1998/1999*, Washington, DC, World Bank, ch. 3.

— (2000), *Global Development Finance 2000*, Washington, DC, World Bank.

Wyplosz, C. (2002), 'How Risky is Financial Liberalization in the Developing Countries?', *Comparative Economic Studies*, 44(2/3), Summer, 1–26.

垄断竞争理论

垄断竞争与最优产品多样性*

在福利经济学中，生产的基本问题是一个市场机制能否导致社会最优的产品种类和产品数量。众所周知，由于以下三个方面的原因，市场有可能会出现问题：分配的公正性、外部性和规模经济。本文着重关注最后一个原因。

基本原则是非常清楚的。[1] 只要生产某产品的成本能够被收益和一个适当定义的消费者剩余所抵消，那么就可以生产该商品。最优的产量取决于什么时候商品的需求价格等于边际成本。如果市场能够做到完全歧视定价，那么这样的最优是可以由市场来实现的。否则，我们就会遇到问题。满足边际条件的竞争性市场是不可持续的，因为它的总利润将为负。加入垄断的因素可以产生正利润，但是同时也违反了边际条件。[2] 所以我们认为，市场是一个次优的解决方案。但是，如果我们要理解偏见的本质，我们必须对该问题有一个更精确的结构化分析。

将这个问题考虑为一个数量对种类的问题是有益的。在规模经济的条件下，通过生产更少种类和更多数量的产品是有利于节省资源的。但是，这将

---

* “Monopolistic Competition and Optimum Product Diversity”, with Avinash K. Dixit, The American Economic Review, Vol. 67, No. 3 (Jun., 1977), pp. 297－308. 斯蒂格利茨的研究部分由斯坦福大学社会科学数学研究所的 NSF SOC74－22182 号科研津贴提供资助。我们要感谢 Michael Spence，感谢审稿人以及对本文早期草稿的评论与建议的执行编辑。

1 关于这一点也见迈克尔·斯彭斯（Michael Spence）的分析。

2 彼德·戴梦德（Peter Diamond）和丹尼尔·麦克法登（Daniel McFadden）对此做了一个简单分析。

导致产品种类的减少，从而造成福利损失。规模经济可以假设，每一种潜在的商品都包含一个固定的建立成本和一个不变的边际成本，从而易于且符合实际的模型化。但是模型化对多样性的需求被认为是很困难的，一些间接的模型化方法曾被采用过。霍太林（Hotelling）的空间模型，兰卡斯特（Lancaster）的产品特征方法，以及均值—方差组合选择模型都曾被采用过。[3] 这些方法导致了包含运输成本或者商品与证券之间的相关关系的结果，从而很难在一般的环境中分析其含义。因此，我们应注意到，传统意义上，定义在所有潜在商品消费数量集上的效用函数无差异曲面的凸性，已经体现了多样性的需求，我们这里采用一种直接的方法去刻画对多样性的需求。所以，一个对消费束（1，0）和（0，1）无差异的消费者将偏好于消费束（1/2，1/2）。这种方法的优点在于，它的结果涉及我们所熟悉的需求函数的自弹性和相对弹性，从而易于理解。

这里，我们着重研究一种非常有意思的情况。这就是，在该种情况下处在同一组或者同一部门，或者同一产业的商品相互之间是良好的替代物，但是相对经济中的其他产品就只有很差的替代性。然后，在这种情况下，我们检验市场机制与最优之间的关系，这种最优既是关于同一组内商品之间的最优，也是关于该组商品与经济中其他商品之间的偏好的最优。我们估计问题的答案既取决于同一组内的替代弹性，也取决于不同组之间的替代弹性。为了尽可能简单地说明这一点，我们将经济的其余部分加总一种商品，标记为0，并把它看作为本位商品。整个经济关于该商品的禀赋被标准化为1；这可以被理解为由消费者所支配的时间。

相关产品的潜在可变范围被标为1，2，3，…分别计各种商品的数量为 x_0 和 $x=(x_1, x_2, x_3, \cdots)$，我们假设效用函数为一个具有凸的无差异面的可分效用函数：

$$u = U(x_0, V(x_1, x_2, x_3, \cdots)) \tag{1}$$

在第一和第二部分，我们将进一步简化假设 V 是一个对称函数，而且所有该组中的商品都具有相等的固定成本和边际成本。那么，此时每一种商品的具体标号就无关紧要了，虽然总的生产的产品种类数 n 仍然是有关的。所以我们可以将那些生产的产品种类标为1，2，…，n，而潜在产品 $(n+1)$，$(n+2)$，…就是没有被生产的产品。这其实是一个具有限制的假设，因为在这样的问题中，我们经常会遇到由于物理上的不同程度的差异而

3 见 Harold Hotelling，Nicholas Stern，Kelvin Lancaster 和 Stiglitz 的文章。

造成的天然的非对称性，性质相近的商品是彼此之间较好的替代品，而性质差别较大的商品彼此之间就不具有很好的替代性。但是，即使没有对称性，我们也能够得到较有意思的结果。在第三部分我们将分析不对称性的一些方面的性质。

同时，我们也假设所有的商品都具有单位收入弹性。这一点与迈克尔·斯彭斯最近的一个研究有所不同。他假设 U 关于 x_0 为线性，从而该产业也服从于局部均衡分析。我们的处理方法使得对跨部门间替代的更好的处理成为可能，但是其他的结果都和斯彭斯的结果很相似。

我们考虑式（1）的两个特殊情况。在第一部分，V 被假设为具有常替代弹性的形式，但是允许 U 有任意形式。在第二部分 U 被假设为具有柯布—道格拉斯形式，但是 V 被假设为具有更一般的可加形式。所以前一种假设允许更一般的跨部门之间的关系，而后一种假设更注重部门内部的替代关系，各自侧重于不同的结果。

收入的分配问题就被忽略了。因此 U 可以被看作代表了萨缪尔森社会无差异曲线，或者（假设适当的加总条件能够得到满足）是被看作一个有代表性的消费者的效用函数的倍数。此时产品的多样性要么可以被理解为不同的消费者消费不同种类的产品，要么可以被理解为每一位消费者的多样性消费。

一、常替代弹性的情况

1. 需求函数

在这一部分的效用函数形式为：

$$u = U\left(x_0, \left\{\sum_i x_i^{\rho}\right\}^{1/\rho}\right) \tag{2}$$

为了凹性条件，我们需要假设 $\rho < 1$。更进一步地，既然我们需要允许一些 x_i 为零的情况，我们需要 $\rho > 0$。我们同样也假设 U 关于其自变量满足齐次性。

预算约束为：

$$x_0 + \sum_{i=1}^{n} p_i x_i = I \tag{3}$$

这里 p_i 为所生产的产品的价格，而 I 为以本位商品表示的总收入，也就是被单位化为 1 的禀赋，再加上公司所获得的利润被分配给了消费者，或者根据情况需要，也有可能要减去为抵消损失所需要的一次性支付。

在这种情况下，一个两阶段的预算过程是合乎情理的。[4] 因此我们定义对偶的数量和价格向量：

$$y = \{\sum_{i=1}^{n} x_i^{\rho}\}^{1/\rho}, q = \{\sum_{i=1}^{n} p_i^{-1/\beta}\}^{-\beta} \tag{4}$$

这里 $\beta = (1-\rho)/\rho$，由于 $0<\rho<1$，所以 β 其值为正。那么可以证明，[5] 在第一阶段：

$$y = I\frac{s(q)}{q}, x_0 = I(1-s(q)) \tag{5}$$

这里函数 s 取决于函数 U 的形式。记 x_0 与 y 之间的替代弹性为 $\sigma(q)$，我们定义 $\theta(q)$ 为函数 s 的弹性，即 $qs'(q)/s(q)$。这时我们有：

$$\theta(q) = \{1-\sigma(q)\}\{1-s(q)\} < 1 \tag{6}$$

这里 $\theta(q)$ 可以为负，因为 $\sigma(q)$ 可以超过 1。

下面转入该问题的第二阶段，容易证明对每一个 i，

$$x_i = y\left[\frac{q}{p_i}\right]^{1/(1-\rho)} \tag{7}$$

这里 y 由式（4）定义。考虑一个单独的 p_i 的变化的作用。这种变化既直接影响 x_i，也通过 q 影响 x_i；从而也通过 y 影响 x_i。现在根据式（4）我们有弹性

$$\frac{\partial \log q}{\partial \log p_i} = \left(\frac{q}{p_i}\right)^{1/\beta} \tag{8}$$

只要该组中产品价格大小的顺序保持不变，这就是（$1/n$）阶的。我们将假设 n 足够大，而且，相应地，我们忽略了 p_i 对 q 的影响；从而也就忽略了对 x_i 的间接影响。这就使得弹性为

$$\frac{\partial \log x_i}{\partial \log p_i} = \frac{-1}{(1-\rho)} = \frac{-(1+\beta)}{\beta} \tag{9}$$

在张伯伦的方法中，这就是 dd 曲线的弹性，dd 曲线是描述在给定其他所有产品的价格情况下，对每一种产品的需求量与该产品价格的关系的曲线。

在我们的组内产品种类较多的情况下，我们也能看到，对于 $i \neq j$，交叉弹性 $\partial \log x_i / \partial \log p_j$ 是可以被忽略不计的。但是当该组中所有的商品价格同时变动，那么个体的较小的作用就变成了一个很大的作用。这就对应于张伯伦

4 见 John Green，p. 21。

5 为了节省笔墨，这些细节以及其他一些内容就被省略了，但是都可以在参考文献中所提到的作者的工作论文中找到。

的 DD 曲线。考虑一个对于所有的 1 到 n，$x_i = x$ 且 $p_i = p$ 的对称情况。我们有：

$$y = xn^{1/\rho} = xn^{1+\beta} \tag{10}$$
$$q = pn^{-\beta} = pn^{-(1-\rho)/\rho}$$

而且根据式（5）和式（7），

$$x = \frac{Is(q)}{pn} \tag{11}$$

此时的弹性是很容易计算的；我们发现

$$\frac{\partial \log x}{\partial \log p} = -\left[1 - \theta(q)\right] \tag{12}$$

此时式（6）说明，DD 曲线向下倾斜。传统的 dd 曲线更有弹性的条件可以由式（9）和式（12）得到

$$\frac{1}{\beta} + \theta(q) > 0 \tag{13}$$

最后我们观察到，对于 $i \neq j$，有

$$\frac{x_i}{x_j} = \left[\frac{p_j}{p_i}\right]^{1/(1-\rho)} \tag{14}$$

所以 $1/(1-\rho)$ 为同一组中任意两种产品间的替代弹性。

2. 市场均衡

我们可以证明任何一种产品都是由一家公司生产。每一家公司都追求其最大化利润，直到边际的进入者只能获得零利润时市场进入才会终止。所以，我们的市场均衡就是人们较为熟悉的张伯伦垄断竞争中的均衡，其中关于数量与多样性的问题总是被提及。[6] 以前的分析未能明确地考虑多样性的需求，而且也忽略了需求的多种组内和组间的相互作用。结果许多经济学家都采用了更含糊不清的假设，认为该均衡导致过度的多样性。我们的分析将挑战其中的某些思想。

每一家公司按照利润最大化行动的条件就是我们所熟悉的边际收益等于边际成本。记 c 为共同的边际成本，注意到对每一家公司的需求弹性为 $(1+\beta)/\beta$，从而对每一家运行中的公司，我们有：

$$p_i\left(1 - \frac{\beta}{1+\beta}\right) = c$$

6　见 Edwin Chamberlin，Nicholas Kaldor 和 Robert Bishop 的著作。

记每一种产品的共同均衡价格为 p_e，我们有：

$$p_e = c(1+\beta) = \frac{c}{\rho} \tag{15}$$

均衡的第二个条件为，企业会一直进入，直到下一个进入者将会赔本。如果 n 充分的大，使得1的增量是一个小量，我们可以假设边际的进入企业恰好持平，也就是说，$(p_n - c)x_n = a$，这里 x_n 由需求函数得到，而 a 为固定成本。由于对称性，该隐含的零利润水平对所有边际内的企业也适用。从而 I 等于1，而且根据式（11）和式（15），我们写出条件，得到经营中企业数目为 n_e：

$$\frac{s(p_e n_e^{-\beta})}{p_e n_e} = \frac{a}{\beta c} \tag{16}$$

给定 $s(p_e n^{-\beta})/p_e n$ 为关于 n 的一个单调函数，那么均衡将是唯一的。这就与我们前面关于两条需求曲线的讨论联系起来。根据式（11），我们可以看到，当 n 增加时函数 $s(pn^{-\beta})/pn$ 的变动就能够告诉我们，当企业的数目增加时，每一家企业所面临的DD需求曲线如何变动。假设它向左平移是很自然的，也就是说，当对每一个给定的 p 值 n 增加时，上述函数值会变小。容易得到关于这一点的条件写成弹性形式为：

$$1 + \beta\theta(q) > 0 \tag{17}$$

这实质上与式（13）的含义是相同的，式（13）描述了dd曲线比DD曲线更有弹性的条件，在后面我们将假设该条件是成立的。

如果 $\sigma(q)$ 比1大得太多，那么该条件不成立。那样的话，n 的增加将使 q 减少，如果需求朝垄断部门的移动程度足够大，就会使得对每一家企业的需求曲线都会向右平移。但是这种情况是不太合理的。

传统的张伯伦分析方法假设，对于作为一个整体的同一组企业有一固定的需求函数。这就相当于假设 $n \cdot x$ 不随 n 的变化而变化，也就是说 $s(pn^{-\beta})$ 不随 n 的变化而变化。而这只有当 $\beta = 0$ 或者当对于所有的 q，$\sigma(q) = 1$ 时才会成立。前者相当于假设 $\rho = 1$，此时同一组中的产品互为完全替代，也就是说，多样性完全没有价值。这将与我们整个分析的目的矛盾。所以，传统的分析隐含地假设 $\sigma(q) = 1$。这就为垄断竞争部门给出了一个不变的预算份额。注意到，在我们的参数化分析方法中，这就意味着DD曲线拥有单位弹性，式（17）成立，因而有唯一均衡。

最后，根据式（7）、式（11）和式（16），我们能够计算出市场中每一家经营中的企业的均衡产出水平：

$$x_e = \frac{a}{\beta c} \tag{18}$$

我们也可以写出，每一组企业作为一个整体的预算份额的表达式：

$$s_e = s(q_e) \tag{19}$$

这里 $q_e = p_e n_e^{-\beta}$。这个结果在后面的比较中我们将要用到。

3. 约束最优

我们的下一步工作是将上述均衡与一个社会最优均衡进行比较。在规模经济条件下，最优或者无约束（实际上只受到技术和资源情况的约束）最优要求定价低于平均成本，因而需要对企业进行一次性转移支付以弥补损失。这样做在概念上和实际中的困难显然很大。因此对于最优，约束最优看来是更合适的。在该约束最优下，每一家企业都赚取非负的利润。这可以通过管制、征税或者免税，或者补贴等手段达到。一个重要的限制就是，一次性转移支付补贴是不可获得的。

我们从这样一个约束最优开始。我们的目的是选择 n，p_i 和 x_i，以便实现效用最大化，满足需求函数，以及保持每一家企业在利润水平均非负的条件。如果所有经营中的公司都必须保持同样的产出水平、价格水平，以及应当严格地赚取零利润，可以在一定程度上简化上述问题。我们就省略了证明过程。这样我们就可以令 I 等于 1，然后根据式（5）效用表成只与 q 有关的函数。当然这是一个递减的函数。从而最大化 u 的问题就变成了最小化 q 的问题，也就是说

$$\min_{n,p} pn^{-\beta}$$

使得

$$(p-c)\frac{s(pn^{-\beta})}{pn} = a \tag{20}$$

为了解这个问题，我们沿着目标函数的一条水平曲线计算对数边际替代率，以及沿着约束曲线的边际转换率，然后令这两者相等。这就得到了条件：

$$\frac{\frac{c}{p-c}+\theta(q)}{1+\beta\theta(q)} = \frac{1}{\beta} \tag{21}$$

二阶条件可以被证明是满足的，简化式（21），我们就可以得到对在该约束最优中生产的所有产品的价格水平 p_c，为

$$p_c = c(1+\beta) \tag{22}$$

对照式（15）和式（22），我们可以看到两种解法得到同样的价格水平。既

然它们面临同样的持平约束，那么也就有同样的企业数目，从而其他所有的变量都可以从这两个量得到。所以这里我们就得到了一个相当令人惊讶的结果，那就是垄断竞争均衡与不存在一次性转移支付条件下的约束最优均衡是完全相同的。张伯伦曾经建议说，这样的一个均衡是“一种理想化的情况”，我们的分析表明，什么时候和在何种意义上这可以变成现实。

4. 无约束最优

我们可以把上述解再与无约束最优均衡或者最优均衡相对照。关于凸性的考虑再一次地说明，所有运营中的企业都必须有同样的产出水平。所以我们将选择 n 家企业，每家产量 x，并使下式最大化

$$u = U(1 - n(a + cx), xn^{1+\beta}) \tag{23}$$

这里我们用到了经济的资源平衡条件以及式（10）。一阶条件为

$$-ncU_0 + n^{1+\beta}U_y = 0 \tag{24}$$

$$-(a + cx)U_0 + (1+\beta)xn^{\beta}U_y = 0 \tag{25}$$

根据第一阶段的预算问题，我们知道 $q = U_y/U_0$。根据式（24）与式（10），我们得到在无约束最优均衡处，每一家运营中的企业所收取的价格水平 p_u，为

$$p_u = c \tag{26}$$

这个结果当然是意料之中的。同时根据一阶条件，我们有

$$x_u = \frac{a}{c\beta} \tag{27}$$

最后，根据式（26），每一家企业刚好完全收回其可变生产成本。那么对企业的一次性转移支付为 an，从而 $I = 1 - an$，且

$$x = (1 - an)\frac{s(pn^{-\beta})}{pn}$$

企业的数目 n_u 由下式定义

$$\frac{s(cn_u^{-\beta})}{n_u} = \frac{a/\beta}{1 - an_u} \tag{28}$$

现在，我们可以比较无约束均衡和受限制均衡情况下对应变量的大小了。让人印象最为深刻的结果就是，在两种情况下每家运营中的企业的产出水平是相同的。在张伯伦均衡中，每一家企业都在最低平均成本点以左经营的事实，传统上被描述为经济中存在过剩的生产能力。但是，当产品多样性同样能增加人们的效用时，也就是说，当不同种类的产品并不是相互间地完全替代时，在一般最优中，并不是将每一家企业的产量提升到所

有的规模经济的好处都被完全利用的水平。[7]在一个并不是极端的情况中我们已经证明了，在最优均衡中，对规模经济的利用程度并没有超出该均衡中对规模经济的利用程度。此时我们可以很容易想像，该均衡对规模经济的利用程度远远超出了从社会角度来看的最优程度的情形。因此我们的结果是从无约束最优和约束最优的角度而言，对过剩生产能力说法的正确性给出了否定的结论。

直接比较式（16）和式（28）所表明的企业数目是比较困难的，但是有一个简单的方法被证明是很简单明了的。显然，比较无约束最优，约束最优均衡有更低的效用水平。同样，其总量收入也要低于前者中的总量收入水平。所以我们必须有

$$q_u < q_c = q_e \tag{29}$$

更进一步地，该差别还必须足够大，使得在无约束最优均衡中关于 x_0 和产量指数 y 的预算约束，落在约束最优均衡情况下相应区域的外面，如图 1 所示。令 C 点为约束最优，而 A 点为无约束最优，且令 B 点为从原点到 C 点的连线与无约束情况下的无差异曲线的交点。根据齐次性，经过 B 点的无差异曲线与经过 C 点的无差异曲线平行。所以每一个从 C 点到 B 点的移动，然后从 B 点到 A 点的移动都会增加 y 值。既然在两个最优处 x 值相同，那么我们必须有

$$n_u > n_c = n_e \tag{30}$$

所以，相对于有约束最优和均衡，无约束最优实际上允许更大的产品多样性；这是另一个与传统的关于过度的产品多样性的说法相矛盾的地方。

根据式（29），我们可以很容易地比较预算份额。在我们所使用的符号系统中，我们发现当 $\theta(q) \geqslant 0$，即，$\sigma(q) \geqslant 1$ 在整个 q 的相关变动范围内都成立时，$s_u \geqslant s_c$。

关于两种情况下的 x_0 的相对大小，不大可能有一个一般结果，仔细观察图 1 可以得到这一点。但是我们可以得到一个充分条件：

$$x_{0u} = (1 - an_u)(1 - s_u) < 1 - s_u \leqslant 1 - s_c = x_{0c} \text{ 如果 } \sigma(q) \geqslant 1$$

在这种情况下，约束最优或者均衡将比无约束均衡使用更多的本位资源。另一方面，如果 $\sigma(q) = 0$，我们就有 L 形的等产量线，在图 1 中，就会有 A 点与 B 点重合，从而得出相反的结论。

在这一部分，我们看到在一个不变的组内产品替代弹性条件下，市场均

7　见 David Starrett 的著作。

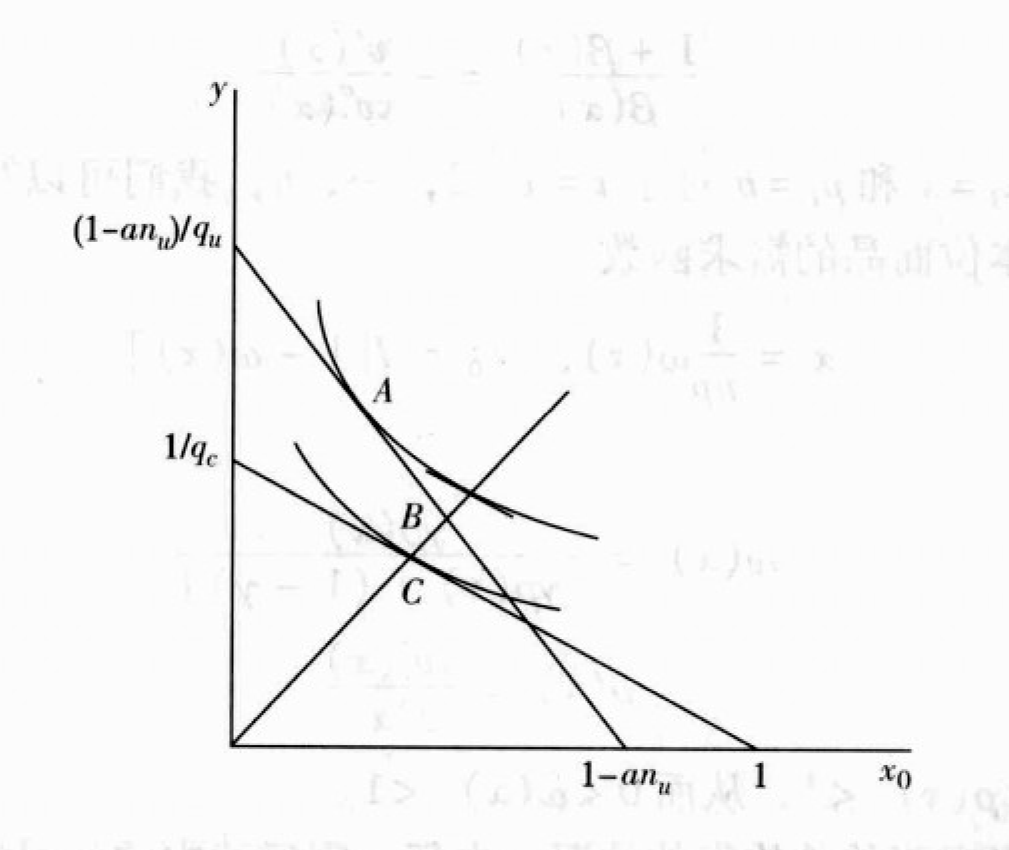


图 1

衡与约束均衡重合。我们也证明了无约束均衡拥有更多的企业，每一家企业都具有相同的规模。最后，部门间的资源分配被证明取决于不同部门间产品的替代弹性。这个弹性也给出了关于均衡唯一性的条件以及最优解的二阶条件。

在本文以后的部分，我们将通过对部门间的替代弹性做出一个特定的假设，以得到分析上的简化。相应地，我们将允许更一般形式的部门内替代弹性。

二、可变弹性的情况

此时的效用函数形式为

$$u = x_0^{1-\gamma}\left\{ \sum_i v(x_i) \right\}^{\gamma} \tag{31}$$

该函数关于 v 递增且为凹函数，$0<\gamma<1$。这就在某种意义上相当于假设单位部门间替代弹性。但是这个假设并不严格，因为组效用函数 $V(\underline{x}) = \sum_i v(x_i)$ 并不是满足齐次性，因此两阶段预算法不再可行。可以证明，在大组情况下，对任意的 i，dd 曲线的弹性为

$$-\frac{\partial \log x_i}{\partial \log p_i} = -\frac{v'(x_i)}{x_i v''(x_i)} \tag{32}$$

这与第一部分的作为 x_i 的函数的情况有所不同。为了突出相同之处与不同之处，我们定义 $\beta(x)$ 为

$$\frac{1+\beta(x)}{\beta(x)} = -\frac{v'(x)}{xv''(x)} \tag{33}$$

下一步，令 $x_i = x$ 和 $p_i = p$ 对于 $i = 1, 2, \cdots, n$，我们可以写出 DD 曲线的方程以及对本位商品的需求函数

$$x = \frac{1}{np}\omega(x), \quad x_0 = I[1-\omega(x)] \tag{34}$$

这里

$$\omega(x) = \frac{\gamma\rho(x)}{[\gamma\rho(x)+(1-\gamma)]} \tag{35}$$

$$\rho(x) = \frac{xv'(x)}{v(x)}$$

我们假设 $0<\rho(x)<1$，从而 $0<\omega(x)<1$。

现在，考虑张伯伦均衡的情况。由每一家运营中的公司的利润最大化条件可以得到，作为共同的均衡产出水平函数的均衡价格：

$$p_e = c[1+\beta(x_e)] \tag{36}$$

注意上式与式（15）相似。将式（36）代入零利润条件我们得到，x_e 由下式决定：

$$\frac{cx_e}{a+cx_e} = \frac{1}{1+\beta(x_e)} \tag{37}$$

最后，均衡企业数可由 DD 曲线和持平条件得到

$$n_e = \frac{\omega(x_e)}{a+cx_e} \tag{38}$$

对于均衡存在的唯一性，我们再一次地使用 dd 曲线比 DD 曲线更富有弹性的条件，以及使 DD 曲线向左平移的条件。但是这些条件都非常复杂和晦涩，所以我们进行了省略。

现在，我们转入约束最优。我们期望在式（34）和持平条件 $px = a + cx$ 的约束下选择 n 和 x，以便让 u 最大化。通过代入，我们可以将 u 表示成为只与 x 有关的函数：

$$u = \gamma^{\gamma}(1-\gamma)^{(1-\gamma)}\frac{\left[\frac{\rho(x)v(x)}{a+cx}\right]^{\gamma}}{\gamma\rho(x)+(1-\gamma)} \tag{39}$$

一阶条件决定了 x_c，有

$$\frac{cx_c}{a+cx_c} = \frac{1}{1+\beta(x_c)} - \frac{\omega(x_c)x_c\rho'(x_c)}{\gamma\rho(x_c)} \tag{40}$$

比较该式与式（37），再根据二阶条件，我们可以证明，给定$\rho'(x)$对所有的x符号保持不变，则

$$x_c \geqslant x_e \text{当} \rho'(x) \leqslant 0 \tag{41}$$

又由于在两种情况下均为零利润水平，那么点（x_e，p_e）和点（x_c，p_c）落在同一条递减的平均成本曲线上，从而当$x_c \geqslant x_e$时，有

$$p_c \leqslant p_e \tag{42}$$

接下来，我们注意到dd曲线与平均成本曲线相切于点（x_e，p_e），而DD曲线比之更陡峭。考虑$x_c > x_e$的情况。此时，点（x_c，p_c）必须落在DD曲线上比（x_e，p_e）更右边的部分，从而必须对应一个更小的企业数目。而如果$x_c < x_e$，那么我们将得到相反的结果。所以，当$x_c \geqslant x_e$时，有

$$n_c \leqslant n_e \text{当} x_c \geqslant x_e \tag{43}$$

最后，式（41）说明，在这两种情况中都有$\rho(x_c) < \rho(x_e)$。从而$\omega(x_c) < \omega(x_e)$，而且根据式（34）

$$x_{0c} > x_{0e} \tag{44}$$

较小部门间的替代弹性将可能使上述结论反过来，正如在第一部分中那样。

关于上述结果的一个直觉上的原因如下。在我们的组内企业数目充分大的假设下，每一家企业的收入与$xv'(x)$成比例。但是，它的产出对于组效用的贡献又为$v(x)$。两者之比就是$\rho(x)$。所以，如果$\rho'(x) > 0$，那么在边际上，每一家企业会发现在社会最优水平基础上扩张生产更有利可图，从而有$x_e > x_c$。给定利润为零的条件，那么这就导致了更少的企业数目的结果。

注意到，相对大小为效用的弹性，而不是需求的弹性。这两者是有联系的，因为

$$x\frac{\rho'(x)}{\rho(x)} = \frac{1}{1+\beta(x)} - \rho(x) \tag{45}$$

所以，如果在某个区间上$\rho(x)$保持不变，那么$\beta(x)$也一样保持不变，而且我们有$1/(1+\beta) = \rho$，这正是第一部分的情况。但是，如果$\rho(x)$可变，那么我们就得不到$\rho'(x)$与$\beta'(x)$符号之间的联系。所以一般而言需求弹性的可变性与我们的讨论并没有联系。但是，与一类重要的效用函数形式却存在联系。例如，对于$v(x) = (k+mx)^j$，并且$m>0$和$0<j<1$，我们发现$-xv''/v'$和xv'/v正相关。很自然地，现在我们当然会预期随着生产的产品种类的增加，任何两种产品的替代弹性都会增加。在对称的均衡中，这就是边际效用弹性的倒数。那么较高的x将对应于较低的n，从而对应于较

低的替代弹性，较高的 $-xv''/v'$ 和较高的 xv'/v。从而我们将会预期 $\rho'(x)>0$，即该均衡将导致比约束均衡情况下更少的和规模更大的企业。再一次地，这使得人们对关于垄断竞争均衡的过度生产能力和过度产品多样化的传统观点产生了怀疑。

无约束最优化问题即为选择 n 和 x 使下式最大化的问题：

$$u=[nv(x)]^{\gamma}[1-n(a+cx)]^{1-\gamma} \tag{46}$$

容易证明上式的解为

$$p_u=c \tag{47}$$

$$\frac{cx_u}{a+cx_u}=\rho(x_u) \tag{48}$$

$$n_u=\frac{\gamma}{a+cx_u} \tag{49}$$

我们可以使用二阶条件证明：当 $\rho'(x)\gtrless 0$ 时，有

$$x_u \lesseqgtr x_c \tag{50}$$

这就跟式（41）类似，因此也得到了相似的该均衡与无约束均衡下产出水平的比较结果。

无约束最优情况下的价格水平当然是三者当中最低的一个。至于企业数目，我们注意到

$$n_c=\frac{\omega(x_c)}{a+cx_c}<\frac{\gamma}{a+cx_c}$$

从而，我们得到一个单方向的比较：

$$\text{如果 } x_u<x_c\text{，那么 } n_u>n_c \tag{51}$$

对该均衡我们有相似的结果。这些就使得无约束最优同时具有较大的企业规模和较多的企业数目成为可能。这个结果也有其合理性，因为无论如何，无约束最优能最有效地利用资源。

三、不对称的情况

到现在为止，本文的讨论都假设组内的对称性。所以生产的产品种类数是相关的，但是任意一组 n 种产品和其他任何也包含 n 种产品的组是一样好的。下一个最重要的改进就是去掉这个假设。我们容易看到此时组内产品之间的相互作用如何导致偏差。因此，当存在建立成本时，如果生产糖的话，那么对咖啡的需求可能会如此的低，使得对咖啡的生产无利可图。但是这也使得另外一种情况成为可能，那就是对于一对互补性产品，一个进入企业是

可以激励两者都生产的。但是，即使所有的产品都是相互替代的，仍然会存在问题。我们将通过考虑一个从两组产品中选择生产产品的产业，然后来检验选择错误的组别是否可能，来说明这一点。[8]

假设除了本位商品之外还有另外两组商品，这两组商品彼此互为完全替代，且每一组都有一个常弹性子效用函数。更进一步地，我们假设对本位商品有不变的预算份额。所以效用函数形式为：

$$u = x_0^{1-s}\{[\sum_{i_1=1}^{n} x_{i_1}^{\rho_1}]^{1/\rho_1} + [\sum_{i_2=1}^{n_2} x_i^{\rho_2}]^{1/\rho_2}\}_S \tag{52}$$

我们假设在组 i 中的每一家企业都具有固定成本 a_i 和不变边际成本 c_i。

考虑两种类型的均衡，每一种类型的均衡都只生产一组商品。这由以下各式给出

$$\bar{x}_1 = \frac{a_1}{c_1\beta_1}, \bar{x}_2 = 0 \tag{53a}$$

$$\bar{p}_1 = c_1(1+\beta_1)$$

$$\bar{n}_1 = \frac{s\beta_1}{a_1(1+\beta_1)}$$

$$\bar{q}_1 = \bar{p}_1\bar{n}_1^{-\beta_1} = c_1(1+\beta_1)^{1+\beta_1}\left(\frac{a_1}{s}\right)^{\beta_1}$$

$$\bar{u}_1 = s^s(1-s)^{1-s}\bar{q}_1^{-s}$$

$$\bar{x}_2 = \frac{a_2}{c_2\beta_2}, \bar{x}_1 = 0 \tag{53b}$$

$$\bar{p}_2 = c_2(1+\beta_2)$$

$$\bar{n}_2 = \frac{s\beta_2}{a_2(1+\beta_2)}$$

$$\bar{q}_2 = \bar{p}_2\bar{n}_2^{-\beta_2} = c_2(1+\beta_2)^{1+\beta_2}\left(\frac{a_2}{s}\right)^{\beta_2}$$

$$\bar{u}_2 = s^s(1-s)^{1-s}\bar{q}_2^{-s}$$

当且仅当对每一家企业生产另外一组中的商品无利可图时，式（53a）为一个纳什均衡。对这样一种商品的需求函数为

8 对于使用局部均衡分析法的另一种可选方法请见 Spence 的著作。

$$x_2 = \begin{cases} 0 & \text{for} \quad p_2 \geqslant \bar{q}_1 \\ s/p_2 & \text{for} \quad p_2 < \bar{q}_1 \end{cases}$$

从而，我们要求

$$\max_{p_2}(p_2 - c_2)x_2 = s\left(1 - \frac{c_2}{\bar{q}_1}\right) < a_2$$

或者

$$\bar{q}_1 < \frac{sc_2}{s - a_2} \tag{54}$$

相似地，当且仅当

$$\bar{q}_2 < \frac{sc_1}{s - a_1} \tag{55}$$

式（53b）为一个纳什均衡。

现在，考虑最优情况。目标函数与约束条件都是为了使得最优均衡中只有一组商品被生产。所以，假设组 i 中的 n_i 种商品被生产，且每一种的产出水平为 x_i，且以价格水平 p_i 被提供。效用水平由下式给出：

$$u = x_0^{1-s}\{x_1 n_1^{1+\beta_1} + x_2 n_2^{1+\beta_2}\}^s \tag{56}$$

且可获资源的约束为

$$x_0 + n_1(a_1 + c_1 x_1) + n_2(a_2 + c_2 x_2) = 1 \tag{57}$$

给定其他变量的值，在（n_1，n_2）空间中函数 u 的水平曲线凸向原点，而约束条件是线性的。所以最终我们肯定会得到一个角点最优解（至于持平条件，除非两者的 $q_i = p_i n_i^{-\beta_i}$ 值相等，那么对一组中商品的需求将为零，而且在该情况下是不可能避免损失的）。

我们注意到，这里已经构造了一个例子来说明如果正确的组别被选定以后，均衡将不会带来关于约束最优的任何有偏性。因此，要找到约束最优，我们只要看看式（53a）和式（53b）中的 u_i 值，然后判断哪个较大。换而言之，我们需要看哪个 q_i 值较小，然后选择式（53a）和式（53b）中所定义的与其对应的情况（这种均衡有可能是，也有可能不是一个纳什均衡）。

我们用图 2 来说明可能的均衡与最优。给定所有相关的其他参数值，我们从式（53a）和式（53b）中计算（$\bar{q}_1$，$\bar{q}_2$）。然后，式（54）和式（55）告诉我们，是否其中一种或者两者都是可能的均衡，而一个关于 $\bar{q}_1$ 和 $\bar{q}_2$ 值大小的简单比较告诉我们，哪种情况是约束最优。在图 2 中，非负象限被分

成了若干区域，在每一个区域我们有一个均衡与最优的组合与其对应。我们所要做的只是确定在该空间中（$\overline{q}_1$，$\overline{q}_2$）的位置，以得到与这些参数值相对应的结果。更进一步地，我们可以比较与不同的参数值相对应的点的位置，从而进行一些比较静态分析。

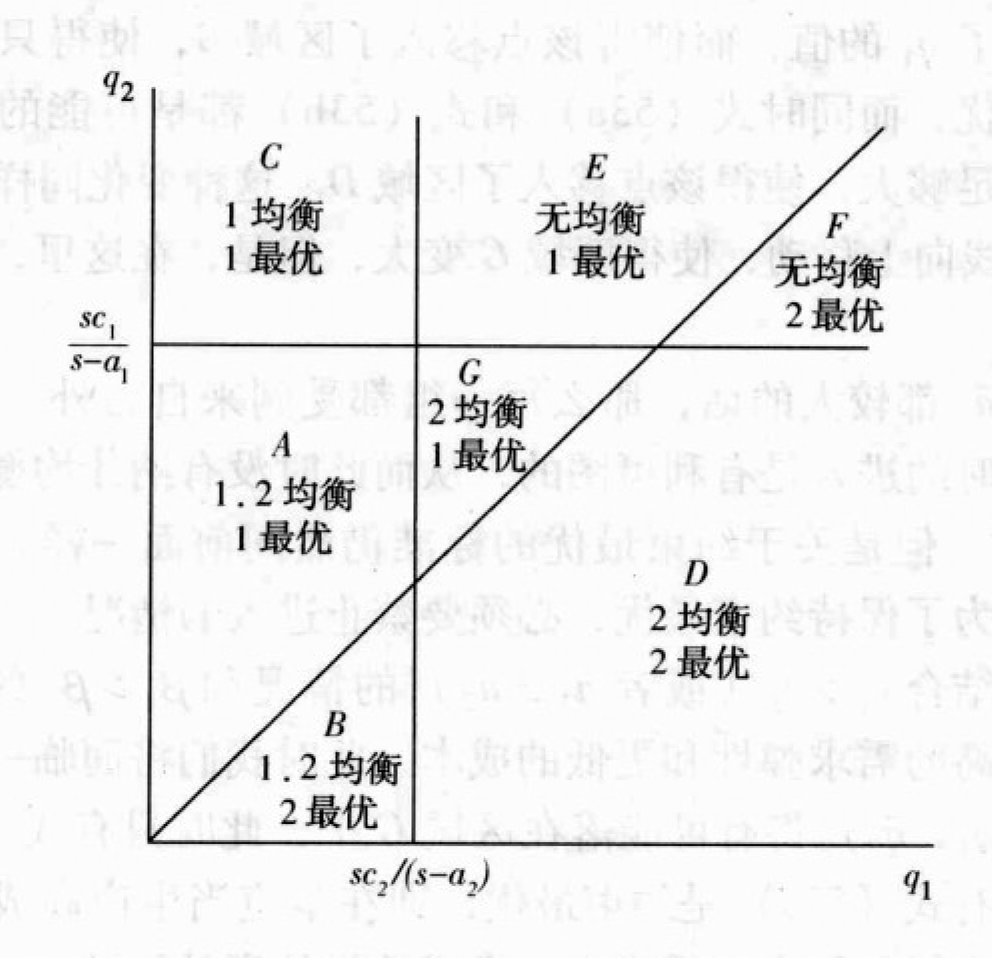


图 2 标记为 1 的是式（53a）的解，标记为 2 的是式（53b）的解

为了理解这个结果，我们必须了解 $\overline{q}_i$ 的值如何取决于参数值。容易看到，两者均为 a_i 和 c_i 的递增函数。我们也发现

$$\frac{\partial \log \overline{q}_i}{\partial \beta_i} = -\log \overline{n}_i \tag{58}$$

而且我们预期该值较大且为负。进一步地，根据式（9），我们知道一个较高的 β_i 值，对应于该组中任意产品需求函数的较低的自价格弹性。所以 $\overline{q}_i$ 为关于该弹性的递增函数。

考虑最初的一个对称情况，$sc_1/(s-a_1)=sc_2/(s-a_2)$，$\beta_1=\beta_2$（从而区域 G 就不存在了），而且假设点（$\overline{q}_1$，$\overline{q}_2$）落在区域 A 和 B 的交界线上。现在考虑一个参数值的变动，比如，对组 2 中产品的一个较高的自弹性。这就提高了 $\overline{q}_2$ 的值，使得该点移入了 A 区域，从而只生产组 1 中产品变为最优。但是式（53a）和式（53b）都是可能的纳什均衡，从而有可能均衡时，在本应该低弹性组中的产品被生产时，高弹性组中的商品被生产了。如果弹性的差别足够大，那么该点将会移入 C 区域，这时式（53b）就

不再是一个纳什均衡。但是由于固定成本的存在，在由组 1 中商品进入对该“错误”均衡的破坏发生威胁之前，弹性的足够大的差别是必要的。相似的结论也适用于区域 B 和 D。

接下来，我们再一次地从对称情况开始，考虑一个较高的 a_1 和 c_1 的情况。这就提高了 $\bar{q}_1$ 的值，而使得该点移入了区域 B，使得只生产低成本组的产品成为最优，而同时式（53a）和式（53b）都是可能的纳什均衡，直到成本的差别足够大，使得该点移入了区域 D。这种变化同样使得区域 A 和 C 之间的分界线向上移动，使得区域 G 变大，但是，在这里，这一点并没有多大意义。

如果 $\bar{q}_1$，$\bar{q}_2$ 都较大的话，那么每一组都受到来自另外一组的进入的威胁，因为，此时的进入是有利可图的，从而此时没有纳什均衡存在，诸如区域 E 和 F 那样。但是关于约束最优的标准仍然同前面一样。因此，我们有可能存在一个为了保持约束最优，必须要禁止进入的情况。

如果我们结合 $c_1 > c_2$（或者 $a_1 > a_2$）的情况和 $\beta_1 > \beta_2$ 的情况，即组 2 中的商品有更高的需求弹性和更低的成本，此时我们将面临一个更坏的可能性。因为点（$\bar{q}_1$，$\bar{q}_2$）将有可能落在区域 G 中，此时只有式（53b）是可能的均衡，而只有式（53a）是约束最优，即在本应当生产高成本，低弹性组商品种类时，市场只能生产低成本，高弹性组的商品种类。

大致说来，关键在于需求函数缺乏弹性的商品有可能赚取超过可变成本的收入，但是它们也有较大的消费者剩余。所以，较之最优情况，市场是偏向于它们还是偏离它们并没有一个直截了当的答案。这里，我们发现答案是后者，而迈克尔·斯彭斯在其他的环境下独立的研究也证实了这个结果。相似的结果适用于边际成本的差异。

在阐释拥有消费者异质性和社会无差异曲线的模型时，需求无弹性的商品将是那些只有一部分消费者有强烈需求的商品。所以，我们将得到一个相对于足球赛，市场对歌剧偏离的“经济”理由，也为在给定收入分配最优的前提下，对歌剧进行补贴和对足球赛征税提供了解释。

如图 3 所示，即使当交叉弹性为零时，也有可能存在对生产的商品的不正确选择（或者相对于无约束最优，或者相对于约束最优）。图 3 描述了一种这样的情况：商品 A 有比商品 B 更富有弹性的需求曲线；在垄断竞争均衡中，A 商品被生产，而 B 商品没有。但是，很明显，生产 B 商品是对社会有利的，因为当忽略了消费者剩余，它刚好处在边际上。所以，那些应当生产而没有被生产的商品，就是那些具有缺乏弹性需求函数的商品。事实上，

正如在关于垄断竞争的传统分析中，去掉一家企业，使得对其他企业的需求曲线向右平移（即增加了对其他企业产品的需求），如果来自产品 A 的消费者剩余（在其最优的产出水平上）小于来自产品 B 的消费者剩余（即在图中阴影区域的面积超过有斑纹区域的面积），那么，约束帕累托最优将导致对具有更富弹性的需求函数的商品的生产的限制。

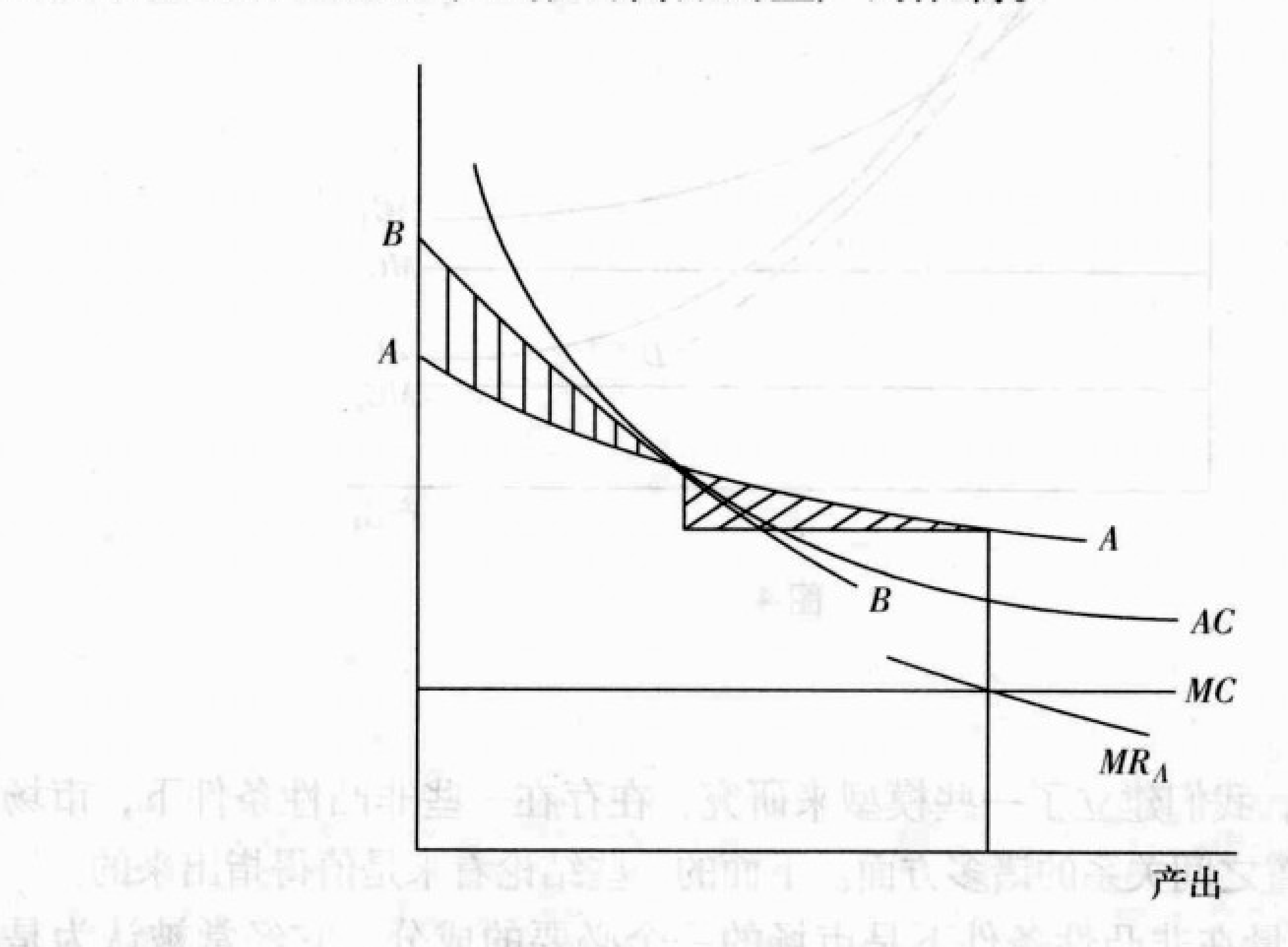


图 3

类似的分析也可以适用于具有相同的需求函数，但有不同成本结构的商品。如果商品 A 被假设有较低的固定成本和较高的边际成本。这样平均成本曲线就会仅相交一次，如图 4 所示。在垄断竞争均衡中商品 A 被生产，而商品 B 没有（虽然它处在被生产的边际上）。但是，再一次地，我们注意到商品 B 应当被生产，因为它有较大的消费者剩余；事实上，如果商品 B 被生产，那么商品 B 将在一个比商品 A 要高得多的产出水平上进行生产，从而有一个大得多的消费者剩余。所以如果政府禁止对商品 A 的生产，那么对商品 B 的生产将是可行的，而且社会福利也将得到增进。

在约束帕累托最优和垄断竞争均衡的比较中，我们可以观察到，在前者中我们用一些高固定成本低边际成本的商品代替了一些低固定成本高边际成本的商品，用一些具有缺乏弹性需求函数的商品代替了一些具有富有弹性需求函数的商品。

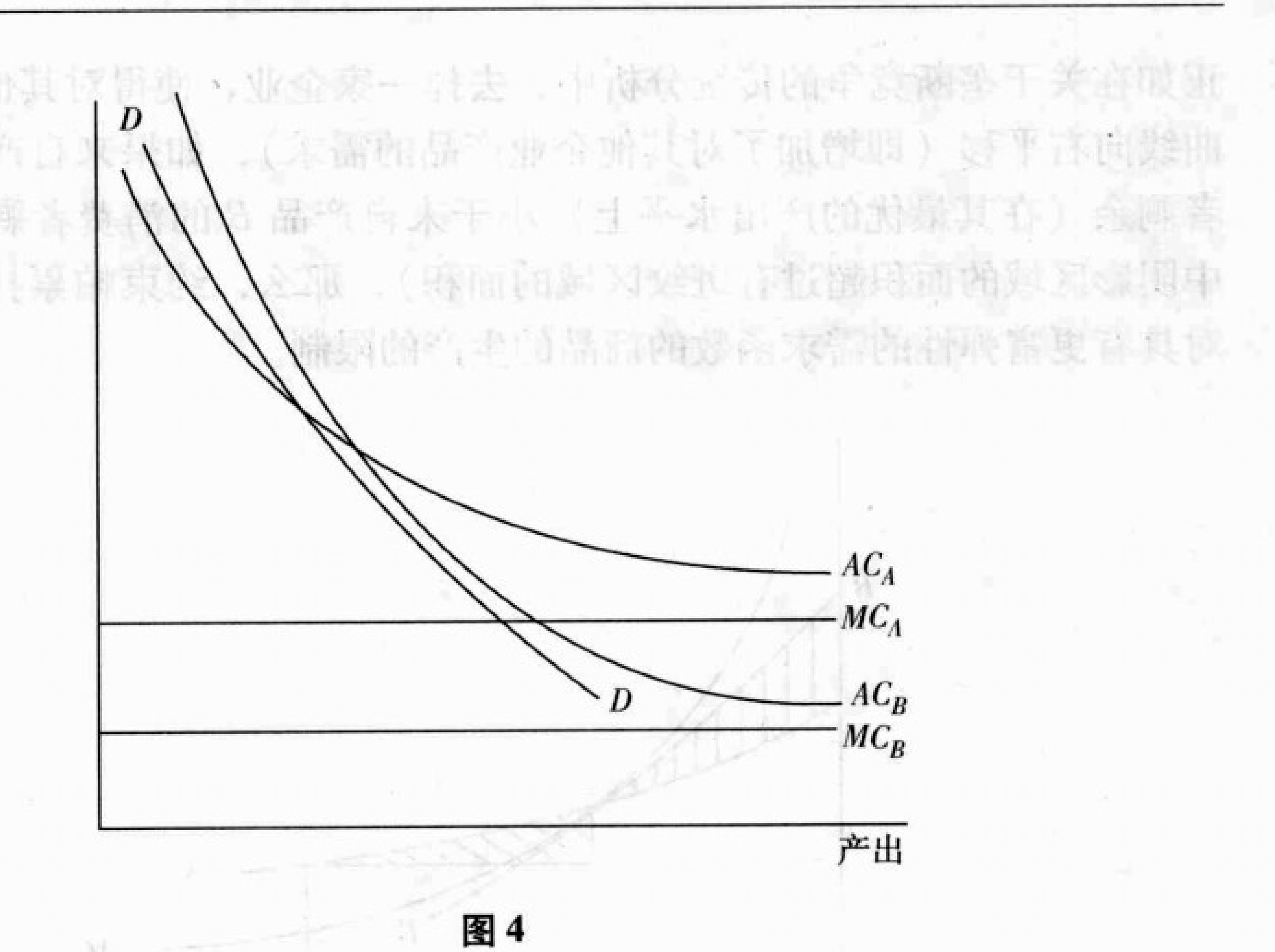


图 4

四、结论

在本文中，我们建立了一些模型来研究，在存在一些非凸性条件下，市场与最优资源配置之间关系的诸多方面。下面的一些结论看来是值得指出来的。

垄断的力量在非凸性条件下是市场的一个必要的成分，它经常被认为是扭曲资源配置，使得资源流出相关部门。但是，在我们的分析当中，垄断的力量使得公司能够支付固定成本，而进入也不能被排除，所以垄断力量与市场扭曲方向的关系不再显而易见。

在我们的常弹性效用函数形式的主要例子中，不管弹性的值为多少（从而不管它所意味着的相应的需求函数的弹性值是多少），市场结果就是受限制的帕累托最优。在可变弹性条件下，两个方向的偏离都是可能的，这不取决于需求弹性的变化，而是取决于效用函数弹性的变化。我们认为，在垄断竞争行业，市场机制将会导致过少的企业数目。

在不对称的需求和成本条件下，我们也得到了对于具有缺乏弹性和高成本的商品种类的偏离。

隐含在这些结果背后的一般原则是，市场机制在合适的边际上考虑利润，而社会最优考虑消费者剩余。但是，对该原则的应用取决于成本和需求函数的具体情况。我们期望这里讨论的一些情况，与我们指出的其他一些研究一道，能够提供一些有用的新见解。

参考文献

R. L. Bishop, "Monopolistic Competition and Welfare Economics," in Robert Kuenne, ed., *Monopolistic Competition Theory*, New York 1967.

E. Chamberlin, "Product Heterogeneity and Public Policy," *Amer. Econ. Rev. Proc.*, May 1950, *40*, 85–92.

P. A. Diamond and D. L. McFadden, "Some Uses of the Expenditure Function In Public Finance," *J. Publ. Econ.*, Feb. 1974, *82*, 1–23.

A. K. Dixit and J. E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," econ. res. pap. no. 64, Univ. Warwick, England 1975.

H. A. John Green, *Aggregation in Economic Analysis*, Princeton 1964.

H. Hotelling, "Stability in Competition," *Econ. J.*, Mar. 1929, *39*, 41–57.

N. Kaldor, "Market Imperfection and Excess Capacity," *Economica*, Feb. 1934, *2*, 33–50.

K. Lancaster, "Socially Optimal Product Diferentiation," *Amer. Econ. Rev.*, Sept. 1975, *65*, 567–85.

A. M. Spence, "Product Selection, Fixed Costs, and Monopolistic Competition," *Rev. Econ. Stud.*, June 1976, *43*, 217–35.

D. A. Starrett, "Principles of Optimal Location in a Large Homogeneous Area," *J. Econ. Theory*, Dec. 1974, *9*, 418–48.

N. H. Stern, "The Optimal Size of Market Areas," *J. Econ. Theory*, Apr. 1972, *4*, 159–73.

J. E. Stiglitz, "Monopolistic Competition in the Capital Market," tech. rep. no. 161, IMSS, Stanford Univ., Feb. 1975.

创新理论

关于技术变革的新观点*

关于技术变革的最新文献几乎都是基于如下假设：技术变革的效应可以由生产函数的向外移动来代表（如图1所示）。技术进步被认为在所有可能的技术条件下，将提高人均产出。然而，这一方法的倡导者似乎已经忘记了新古典生产函数的源头：随着生产过程处理数目的增多（在一个生产活动分析模型中），生产可能性边界线将越来越渐近于平滑、可微的曲线。但是，曲线上不同的点，仍然代表着不同的生产过程，与这种生产技术相联系的是相应程度的特定技术知识。的确，不管是新古典理论的支持者还是批评者，似乎都忽略了一个生产活动分析方法（罗宾逊夫人（Mrs. Robinson）的蓝图）的最重要的方面：如果人们仅仅在蓝图上的某一方面做一定程度的技术改进，这对整个蓝图的其他部分要么作用有限，要么根本不起作用。如果技术进步仅仅是对于生产产品的某一技术而不是所有技术而言，则综合的生产函数的变化，将是生产可能性边界上某一点的向外移动，而不是整个曲线的移动（见图2）。图2表示一个极端的情形，也就是，技术进步是对于某一技术的完全"局部化"（localized）：对其他技术改进没有任何的溢出效

---

* "A New View of Technological Change," with Anthony B. Atkinson, *The Economic Journal*, Vol. 79, No. 315 (Sep., 1969), pp. 573–578.

作者感谢迪蒙尼尔（G. de Menil）、戴梦德（P. A. Diamond）、艾克劳斯（R. S. Eckaus）、哈恩（F. H. Hahn）、皮奥尔（M. Piore）、罗斯切尔德（M. Rothschild）、谢尔（K. Shell）和威廉姆森（J. H. Williamson）等人的对早期草稿的有益评论。斯蒂格利茨的研究部分由美国—英国教育委员会（United States – United Kingdom Educational Commission）资助，部分由美国国家科学基金会资助。

应。在现实中，我们应该预期，特定技术进步将带来溢出效应，并且会影响某些技术。不过，只有当溢出效应对每一*种*技术都产生影响时，我们才能回到传统的立场上。这意味着，技术进步只有是这样，才能提高生产力——譬如说，从每一种最原始的手工纺织机到全自动化纺织机，都有相应的技术进步。[1] 在本文中，我们将检验技术进步“局部化”的某些含义，并将用这一视角与技术进步导致生产函数整体移动的传统假设做比较。

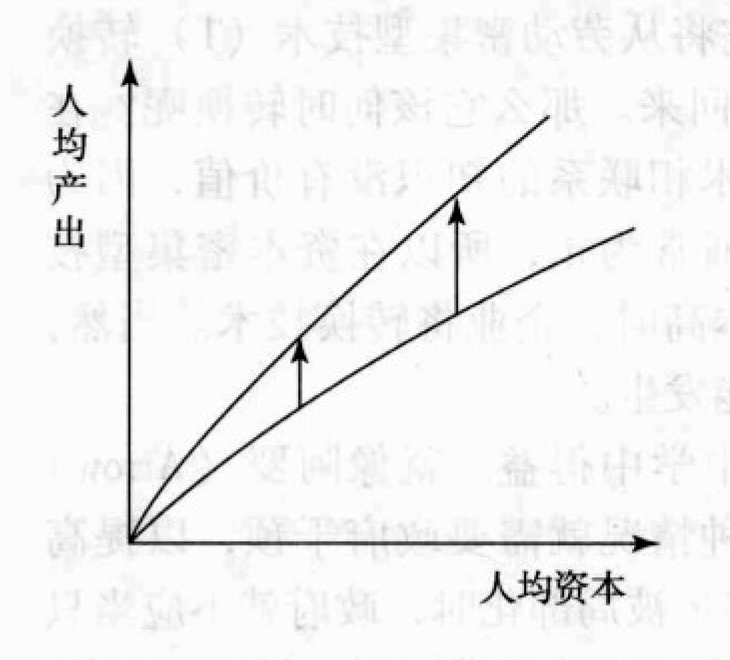


图 1

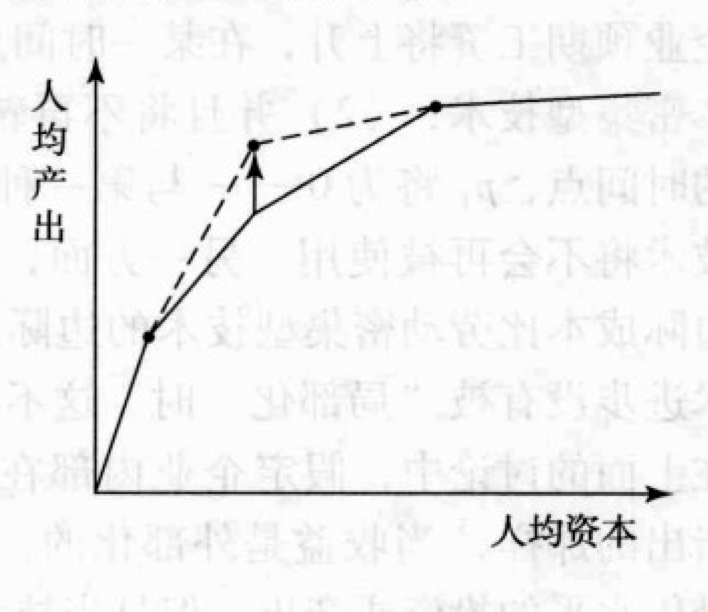


图 2

大体而言，技术知识的增加，要么来自于生产经验（干中学），要么来自于研究活动。在现有文献中，这两个因素都受到相当的关注，但是，当技术进步呈现“局部化”时，有几个重要的含义至今仍然未被揭示。我们先来考虑干中学的情形。如果通过学习而获得的知识被“局部化”，则生产函数的移动，将被固定在企业（或经济体）当前运营的技术点上。这反过来意味着，当一个企业决定使用何种技术时，它必须考虑其当前的技术选择对未来生产可能性的影响。现在来考虑某家企业的情况。该企业的学习被有效地内部化，它在两种技术间选择，一种技术比另一种技术更加资本密集。如果它采纳更具有劳动密集型的技术，这意味着此种技术的生产力将随着学习而增加，而若采纳较资本密集型的技术，则生产力将保持不变。因此，它不能仅仅根据当前的要素价格来决定技术的选择，也必须考虑与每一种技术相联系的知识增加的价值——也就是说，企业不能完全“目光短浅”地行事。成本最小化的条件是：

1 事实上，通常所用的哈罗德中性（Harrod neutrality）假设，假定每单位产出的劳动需求的百分比减少，对所有技术是相同的。

$$C_1 - p_1 = C_2 - p_2$$

这里 C_i 是第 i 种技术的边际成本；p_i 是使用第 i 种技术每多生产 1 单位产品而在知识上收获的价值。如果企业对第二种技术知识增加的价值看得充分高的话，它就将使用这一技术，即使当前的要素价格将导致成本更高（这一论断可以用来证明，诸如采用原子能技术的核电厂的选择合理性：即使每千瓦时的成本更高，但是，在建设过程中的知识获得足以抵消额外的成本）。假设企业预期工资将上升，在某一时间点它将从劳动密集型技术（1）转换至资本密集型技术；（2）并且将不再转变回来。那么它该何时转换呢？在转换的时间点，p_1 将为 0—— 与第一种技术相联系的知识没有价值，因为这种技术将不会再被使用。另一方面，p_2 通常为正，所以在资本密集型技术的边际成本比劳动密集型技术的边际成本高时，企业将转换技术。当然，当技术进步没有被“局部化”时，这不可能发生。[2]

在上面的讨论中，假定企业内部在干中学中得益。就像阿罗（Arrow）已经指出的那样，[3] 当收益是外部化的，这种情况就需要政府干预，以提高社会最优水平的投资或产出。但是当技术进步被局部化时，政府就不应当只是关心投资或产出水平，还应当确保将企业引导向长期考虑下的“正确”技术。这与发展中国家鼓励幼稚产业的行为尤其相关：当技术进步局部化时，政府应该补贴幼稚技术而非幼稚产业。这种情况下，关税可能不仅是低效率的，而且是可能完全无效率的：例如，政府只想鼓励一种“中间式的”制造技术，而不是鼓励生产同一产品的传统，或者高度自动化的方法。

回到技术进步源于研究和创新活动的情况，已经有相当多的文献研究增进技术知识的最优支出数量。当技术进步被局部化于一种技术时，我们要回答另外一个重要问题——我们应该改进哪种技术？研究活动可被引导倾向于任何流程的改进，但一旦它被执行，相应获得的是某一特殊流程的具体知识。以最近的经济增长理论文献的视角来看，技术知识与“水泥—土坯”（putty - clay）资本有相同的特征。这意味着，当企业在选择发展何种技术时，它不仅应该考虑当前的，也必须考虑未来的要素价格。企业没必要将研究支出用于当前使用的技术上，如果企业预期工资上升，将导致去发展和使

2 如果在干中学的情况下，不管技术变革是否被局部化，企业都将在边际收益低于短期边际成本的水平上生产；均衡条件要求边际收益等于短期边际成本减去未来从学习中获得的成本的边际减少的净现值。

3 K. Arrow, “The Economic Implications of Learning by Doing”, Vol. 29, June 1962, pp. 155 - 73.

用资本更为密集型的技术。再次强调，企业不能鼠目寸光。

局部化的技术进步这一概念，对于理解欠发达国家是否应投入资源用于开发新生产技术这类问题，同样有所裨益。

有时，人们会争辩说，将资源配置给研究活动，并不是欠发达国家的当务之急，因为它可从发达国家的技术进步中获益，而任何独立的研究活动只不过是简单地重复劳动而已。不过，正如我们所建议的，如果技术知识对特别生产过程至关重要，情况就不是那样了。当技术进步被“局部化”时，发达国家的技术进步，不管是来自于研究活动或干中学，对根据本国要素禀赋选择资本密集程度较低技术的欠发达国家而言，相对不会有任何影响。事实上，在某些产业中，局部化效应是如此之强，以致先进技术只需相对更少的资本和劳动，就足以超越资本密集程度较低的同类技术。[4] 当技术进步局部化于一种技术，且劳动与资本增加率都为正时，这种情况如何发生就很清楚（见图3）。从图3中，我们也可以很清楚看到，局部化的技术进步，可能导致短期替代弹性的减少。

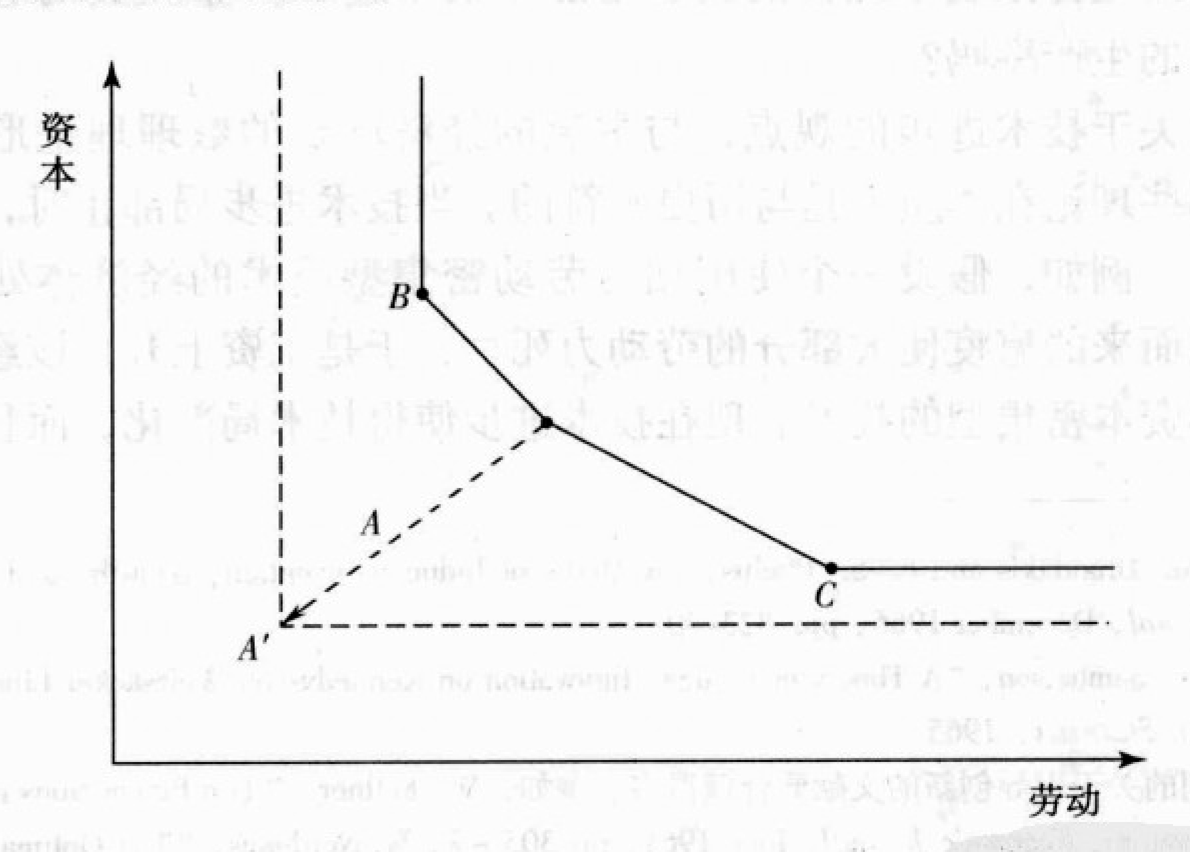


图3

这类先进的资本密集型技术的优越性导致了一种广为流传的观点——发展中国家应该采纳这类先进技术而非劳动密集型技术。不过这一论调并没有

4 关于优越性的例子，请参考 A. K. Sen, The Choice of Techniques (Oxford: Basil Blackwell, 1962); R. S. Eckaus, “The Factor Proportions Problem in Underdeveloped Areas”, *American Economic Review, Supplement*, Vol. L, May 1960, pp. 642-8.

考虑到进行研究的可能性：即使当前资本—劳动比率较低的技术可能是低效的，仍然可以花费资源对它们进行改进。这与最近关于“中间技术”的争论有关。在这一场争论中，人们一直认为，发展中国家应该优先考虑使用人均资本 100 英镑的而不是 1 000 英镑的技术。我们建议，即使这些技术当前尚不存在，发展中国家也应努力去开发。事实上，是否要这么做依赖于一些条件，譬如，先进技术的优势地位，改进传统技术所需的资源，以及计划者的时间表，等等。

我们的分析方法可与引导创新理论的新近版本进行比较，那些理论包含一种创新可能性计划：假设企业面临对不同程度要素增加型技术进步的选择。在这种情况下，当前决策将影响未来生产的可能性，就像技术进步被局部化的情形一样，但是至少采用了两个不合理的假设：（a）在杜朗达基斯—费尔普斯（Drandakis - Phelps）[5] 和萨缪尔森（Samuelson）[6] 模型中，要素增加技术进步方式的选择仅仅是依据当前的要素价格。[7]（b）更为根本的是，企业的选择限制于纯粹的要素增加型技术进步。[8] 企业真的想要提高铲车和推车的生产率吗?[9]

本文关于技术进步的观点，与主流的经济增长的数理理论形成强烈对比，而那些理论在本质上是与历史不符的。当技术进步局部化时，历史是相当重要的。例如，假设一个使用相对劳动密集型技术的经济体处于长期均衡，突然而来的瘟疫使大部分的劳动力死亡，于是工资上升，该经济体开始使用更具资本密集型的技术。现在技术进步使得技术局部化，而且该经济体

---

5 E. M. Drandakis and E. S. Phelps, “A Model of Induced Invention, Growth, and Distribution”, *Economic Journal*, December 1966, pp. 823-40.

6 P. A. Samuelson, “A Theory of Induced Innovation on Kennedy-von Weizsäcker Lines”, *Review of Economics and Statistics*, 1965.

7 早期的关于引导创新的文献要合理得多。譬如，W. Fellner, “Two Propositions in the Theory of Induced Innovation, *Economic Journal*, June 1961, pp. 305 – 8; W. Nordhaus, “The Optimal Rate and Direction of Technical Change”, *Essays on the Theory of Optimal Economic Growth*, K. Shell, ed.（Cambridge: M. I. T. Press, 1967），这些文章在消除不合理假设方面也取得了进步。

8 当然，如果在任何时点上，只有唯一——种技术可供选择，则局部化的问题就变得无足轻重了。C. Kennedy, “Induced Bias in Innovation and the Theory of Distribution”, *Economic Journal*, September 1964, pp. 541-7.

9 Drandakis and Phelps（前引文献）的新近文章，揭示了这一方法将会遇到的困难之处。他们承认，没有理由将创新可能性计划限制于正象限。然而，如果某一要素有“负向的”增加，则新等量曲线将与旧的相交。因此，他们担心，新的等量曲线是否真的是两者的包络线，否则，我们就应该学会忘记这些技术。所有的困难都来源于他们假定的技术进步，对所有流程都发生作用。

很可能将继续使用资本密集型技术，而不是退回到以前的劳动密集型技术。在这一情况下，该经济体的历史在定性和定量上将与没有瘟疫极为不同。例如，在新的长期均衡中资本—产出比率将比旧条件下的更大；如果资本密集型技术中有更多的干中学的话，技术变革的速率也将加快。不过，如果历史是重要的，那么就很有必要以长远的眼光来计划当前的行动。在本文，我们讨论了对企业和对实行计划经济的发展中国家的意义。

产业结构与创新活动的本质*

人们通常把发达工业化国家产出增长的一大部分归因于技术进步。大量的证据表明，这种技术进步并不是以一种随机的方式发生的（例如，Schmookler，1962）。但是能够解释技术进步的速度或者方向的经济学理论实在不多，在有限的能够解释技术进步的理论中，具有很合理的、精确的微观经济学基础的理论，更是少之又少。[1] 这种不足是非常严重的，对于技术进步重要性的认识，使得人们对于用传统的微观经济学模型引导人们去理解现代市场经济的运作和制定政策方针（例如，关于反垄断政策方面）的合理性，产生了严重的质疑。这种不足，同样也是令人惊讶的，因为至少自从熊彼特（Schumpeter，1947）的论文问世以来，发明与创新活动和市场结构是紧密相连的这种思想，已经变成了一种很平凡的理念。而且，肯定会有很多人持有和熊彼特一样的观点，竞争性市场对于发明创造是不利的，而且这种活动所带来的好处，应该要超过由于市场垄断力量所带来的生产无效性所导致的福利损失。[2]

但是，最近关于产业组织理论的很大一部分文献，似乎以一种不同的方

---

* “Industrial Structure and the Nature of Innovative Activity,” with Partha Dasgupta, *The Economic Journal*, Vol. 90, No. 358（Jun. 1980）, pp. 266-293. 这篇文章是1977年8月27日至9月2日在东京举行的国际经济协会关于“*经济增长与资源*”学术大会上所做报告（Dasgupta and Stiglitz, 1977）的第一部分的修改和扩展版本。在当时，达斯古普塔（Dasgupta）教授是Jawaharlal Nehru大学国际事务学院和德里市德里经济学院的访问教授，而斯蒂格利茨是Oskar Morgenstern著名的数理经济学专家，同时在1978年秋季期间，也是普林斯顿大学高级研究中心的访问教授。作者从与Hans Biswinger, Sukhamoy Chakravarty, A. K. Dasgupta, Paul David, Richard Glibert, Sanford Grossman, Glen Loury, Edwin Mansfield 等人的讨论中获益，特别地，作者从 Ashok Guha 对本文的评论中受益匪浅。

1 关于最近的理论，请见 Arrow（1962）, Barzel（1968）, Stiglitz（1970）, Kamien and Schwartz（1972）, Evenson and Kieslev（1975）, Nelson, Winter and Schuette（1976）, Dasgupta and Stiglitz（1977）, Loury（1977）, Nelson and Winter（1977）Dasgupta and Stiglitz（1978）, Levin（1978），以及 Schumpeter（1947）的开创性工作。

2 “一个在任何的给定时点上都在最大程度上利用了其潜能的系统，可能在长期要劣于不在任何给定时间点上这么做的系统，因为也许后者的在这一点上的失败，正是长期表现的速度或者水平的条件”（Schumpeter, 1947, p. 83）。应当注意到在这里熊彼特比较的是产业资本主义和完全竞争。

式在对熊彼特的理论进行检验。[3] 现在人们时常认为产业集中和大规模能够*刺激*发明创新。就好像集中是创新的原因一样。[4]

在本文中，我们尝试为市场结构与创新活动的本质之间的关系提供一个分析框架。要这么做，我们必须对这种新熊彼特观点进行根本的修改。我们将得到，除了在短期中市场结构与创新活动的本质都是内生的，而且就像近期的关于产业组织理论的文献所得到的那样，一个行业的集中度不应该被视做给定的；它们都取决于更基本的因素，比如研究技术、需求条件、资本市场的本质（即市场利率以及公司为了支持研究和开发（R&D）的融资能力），以及法律的结构（例如，专利制度）。为了更加明确，我们将探讨市场集中度与创新活动本质之间的关系。但是，由于它们都是内生变量，那么，不像新熊彼特理论认为的那样，它们之间的关系就不应该被看作是一种因果关系。

我们研究的主要目的是建立一个模型，在这个模型中，可以评估市场经济的有效性，而且短期生产的有效性与长期中的动态收益之间的权衡可以被很严格地讨论。不奇怪的是，一旦这些问题被合适地提出，结果是比第一眼看起来要复杂得多。这不像一家公司做出某种单一决策（比如，总的研发支出的大小）那样简单，而是多家公司做出多种决策的复杂系统的情形；而我们所关注的正是这个复杂系统的结构。比如，每一家公司不但要决定研发方面的支出是多少，而且还要决定采取什么样的研究策略。更进一步地，研究策略可以在下面各个方面有所差异：不但可以在成功到来之日的概率分布方面有所不同，也可以在当研究的主要目标失败时，我们能够学到多少上有所不同，还可以不论在过程上还是在目标上，公司自己所进行的研究与其他公司所从事的研究活动到底有多少相似之处而有所差异。这些决策当中的每一个决策不但会对总的技术进步率，而且会对产业结构和整体经济表现产生重要影响。如果公司倾向于相互抄袭彼此的研究战略，那么大部分的研发支出实际上是重复的，从而导致了社会浪费。如果企业从事风险过高的项

3 实证结果表明，虽然在一定的临界点之前产业集中和创新活动是正相关的，但是，当产业集中程度太高时，这两者之间就是负相关了。关于实证研究的文献综述请见 Scherer（1970），Kamien and Schwartz（1975）。

4 所以，比如，当注意到技术变迁对于市场结构有影响，在 Scherer（1970）的文章中关于 R&D 那一章的主体为：“……考虑到一种可能的反向的因果关系；从市场结构到技术创新”，而且，在其他部分，提出问题“垄断力量，就像市场高度集中的情形那样，真的是适合创新和技术进步生长的土壤吗？”Scherer（1970），p. 347。

目，这就有可能导致技术进步速度过快以及高度的产业集中。而这最终会导致很大的生产效率损失。从公司角度出发的任何一个决策，不得不在内生的产业结构中做出，这一点使我们的分析变得更为复杂。

关于市场结构对研发影响的问题，当然早已不是什么新问题。在竞争和垄断情形下对研发投资水平过低或许是一个一般的假定。这个假定部分地基于知识——研发的产品——具有公共品特征的事实；但是也部分地基于阿罗（1962）的论点，仅仅依赖于成功企业在不同市场结构中盈利大小的比较。

在第一部分，我们将回顾这一分析方法及其错误之处。这种回顾也帮我们为接下来的正式分析建立了平台。事实上本文的一个中心结论是：总而言之没有理由认为市场经济将会保持过低水平的研发投资。它甚至有可能导致过高的研发投资。

第二、三、四部分包括了对于研发活动及其与市场结构的关系等多方面的分析。这些讨论都是基于第一部分中引入的一个关于创新活动的简单模型的扩展模型。在第二部分，我们考察了市场结构与研发投资之间的关系。虽然我们将要建立的模型很简单，但是其含义对于说明我们前面提出的一些要点以及后面将要提出的一些重点却是足够丰富的。在开始前，我们也有必要指出第二部分中的模型的一个中心特征是：虽然从公司的角度来说，研发支出更多的像是生产中的固定成本，但是这些“固定成本”变量的大小本身就是公司的一个选择结果（公司可以从事更多或者更少的研发）。这一点就导致了关于研发竞争分析与传统的关于产品竞争分析的一个本质不同，我们在后续的讨论中将会进行说明。进一步地，我们还应当注意到，在知识的生产和使用过程中，将会表现出一些基本的非凸性质，既然知识是研发投入的产出，那么任何关于研发的分析，必须考虑到这些可能的非凸性。为了简单地说明这一点，应当注意到，一种商品的间接社会收益函数是关于该商品单位生产成本的一个递减但是凸的函数（图4）。这一点的一个直接的含义，就是在两个产出而支出多少相等的研发战略之间，风险规避的社会将会偏好于那一个风险相对较高的战略。我们将在第二部分说明，这一点的另外一个含义就是：即使假定增加的研发投入在降低成本方面具有递减的回报，那么研发支出的净社会收益函数也不一定是一个凹函数。关键在于同样一种知识原则上可以被应用到任何规模的操作中。因此，随着规模的增大，平均每单位规模的信息成本递减；但是每单位规模的信息的价值却并不一定递减。大体来说，不仅是因为知识具有公共物品的特征，还因为在面对诸如道德风险等现象时，建立一个完整的完全竞争的或有市场（Contingent market）集合

基本是不可能的，所以早期的研究着重于价格系统在保持对知识的有效生产和利用方面的不足。我们的部分分析会注意到，知识的生产与利用过程中天然的非凸性质。我们的一般目的就是搞清楚这其中的每一个特征如何影响产业结构以及创新活动的规模和方向。

第二部分中讨论的例子没有涉及时间概念，而且不存在不确定性。它存在一些缺点，其中之一就是不能考察研究活动中的风险承担度，另外一个就是不能确认研发的速度（发明创新活动的节奏）。从而，它也就不能研究这些特征当中的每一点与市场结构具有怎样的联系。当然，事前我们就清楚，每一个特性不但与产品市场结构而且与研发活动的竞争程度相联系。[5] 接下来的模型将能敏锐地抓住这一特征。第三部分讨论的例子与第二部分的相似，但是现在我们假设研究活动具有不确定的收益。在本文中我们关注过程创新（即被设计用来降低生产成本的研发）。这意味着注意力被转向至极端值的分布。正如艾文森和凯斯勒夫（Evenson and Kieslev，1975）所注意到的那样，这又意味着即使风险规避的企业也会愿意从事随机化。第三部分所讨论的例子将证实此点。但是，在我们的模型中，对随机化的需求将会被我们前面已经注意到的特征所加强。

在第四部分的分析中，我们引入了时间。我们假设第一家发明的企业攫取了在所有企业中分配的利润（即我们假设赢家赢得全部）。毫无疑问，这是一个简化的假设。当我们注意到企业经常可以绕过专利进行创新活动，从而最先取得突破的企业，并不一定就是占据最有利位置的企业的事实时，我们应该很清楚朝什么方向修改该模型。

关于第四部分结构的一个关键假设就是：所有的企业都有义务遵守同样的研发战略；也就是说，它们都面临同样的决策树。这一点有很重要的结果。在本文的续文中我们考察了与这种情形相反的极端情形，在那里我们假设企业面临研发成功日期的不确定性，且这些不确定性之间相互独立。

在第五部分中，我们总结了在本文研究中我们认为的最基本的思想。在附录中我们给出了第二、第三部分内容的详细论证，并且对在正文中得到的一些结论给出了严格的证明。特别地，附录I包含了一条可能具有更广泛意义的关于在自由进入条件下，均衡存在性定理。

---

5 在其他几点中，正是这一点最终将本文的分析与阿罗（1962）的分析区分开来，在那里阿罗着重研究不同的产品市场结构，在分析不同市场结构下的创新活动的激励本质时，假设在研发活动中不存在竞争。关于这一点，我们将在第一部分作较大篇幅的分析。

一、过程创新的市场有偏性

在本部分，我们假设不存在收入效应。令 Q 表示一种给定商品的数量。消费 Q 的总的社会效用函数为 $u(Q)$，且 $u'(Q)>0$，$u''(Q)<0$。市场需求由下式给出

$$p = p(Q) = u'(Q) \tag{1}$$

记 $R(Q)=p(Q)Q$ 为一个垄断者的总收入函数。并且假设此时边际收益关于产出递减。

假设现有的生产该商品的最好的技术的单位生产成本为 c。同样也假设一个特定的过程创新将单位生产成本降至 c^*。在阿罗（1962）的开创性论文中，他考察了在三种不同的产品市场结构下创新者的回报。第一种情形就是社会经营市场，在该市场中价格为生产成本。对于这种情形每一期的创新的回报就是净社会剩余的回报，如图 1 中的区域 $ABCD$ 所示。令 π_s 表示这种情况下的收益。第二种情形就是纯粹的垄断市场的情形（即存在进入障碍）。令 π_m 表示如果垄断者进行该项创新所带来的垄断利润的增加。在图 2 中该垄断利润的增加可以由区域 $IJKL$ 和区域 $EFGH$ 之间的差来表示。第三种情形就是竞争性市场的情形。假设 c—成本生产技术已经被竞争性企业广泛地采用了，而创新者要求一个关于 c^*—成本生产技术的专利。令 π_c 表示在专利有效期内每一期创新者所获得的垄断利润。显然，这里有两种情况需

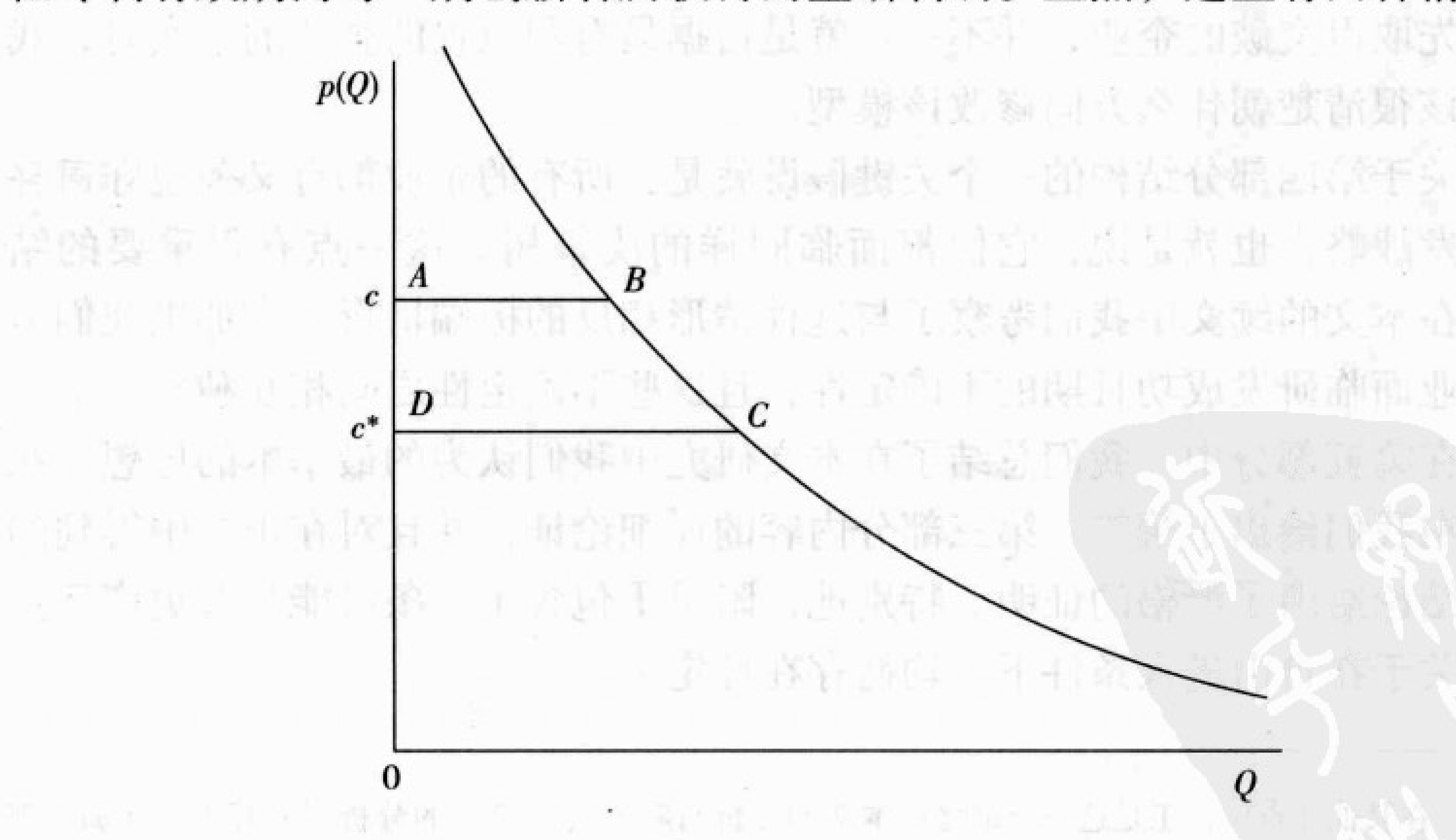


图 1

要考虑。图 3 表示了垄断价格超过 c 的情形。所以这里就存在一个有限价格的现象，创新者将供应整个市场，同时收取 c 的价格。对于这种情况 π_c 由区域 $ABC'D$ 给出。

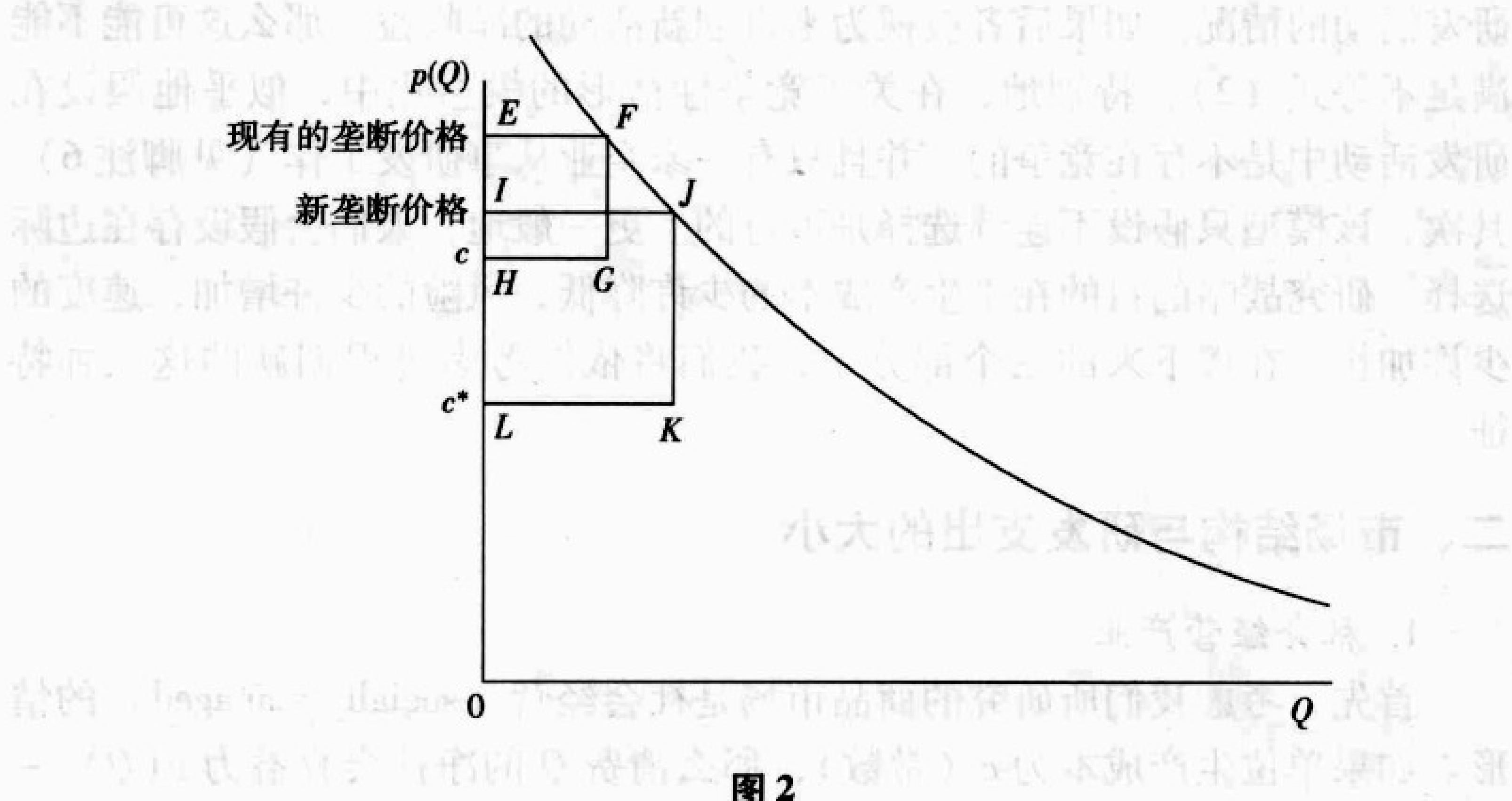


图 2

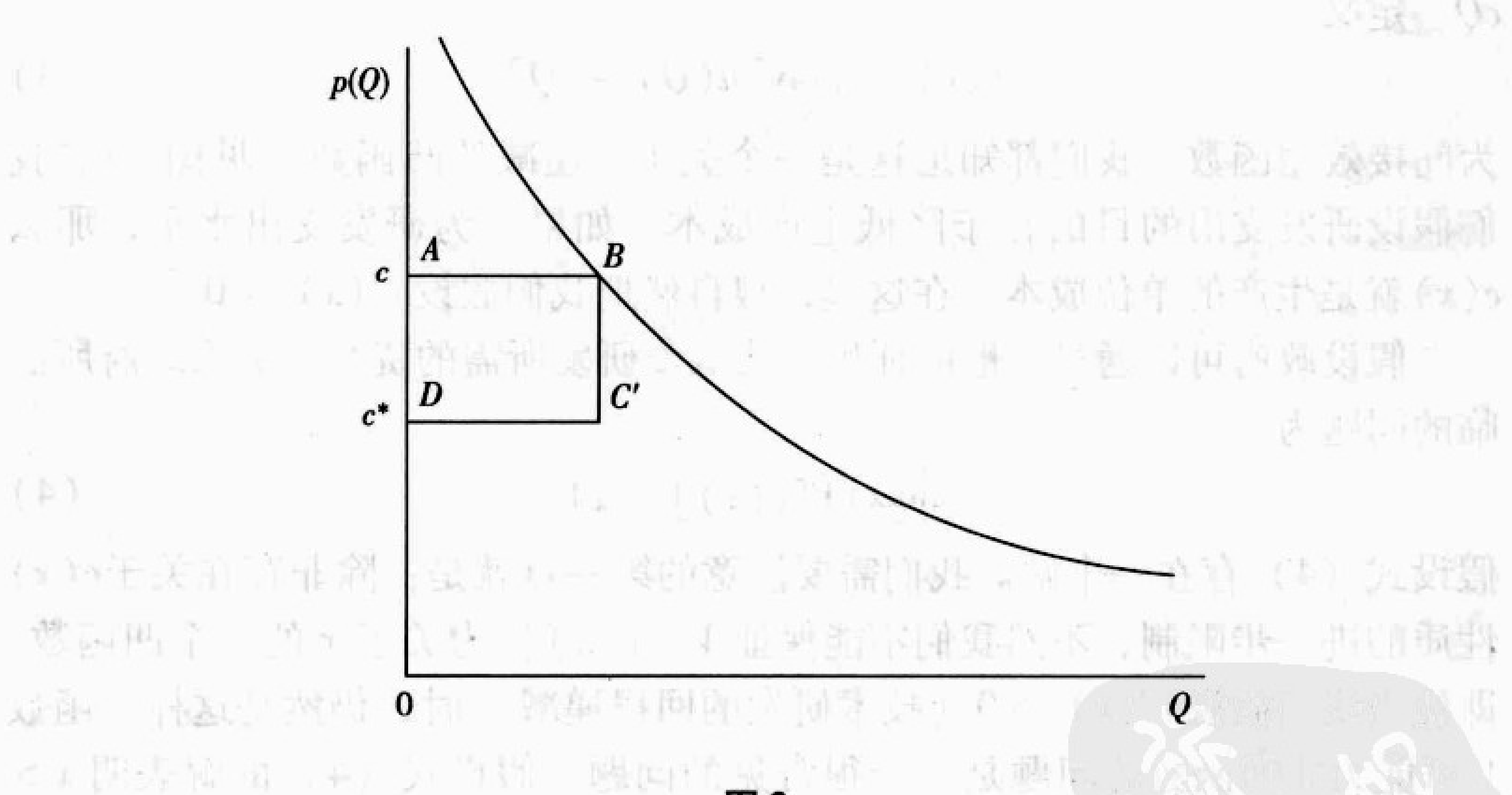


图 3

现在可以证明

$$\pi_s > \pi_c > \pi_m{}^6 \tag{2}$$

通过这个结果，阿罗认为，"在垄断情况下的创新激励要小于在竞争情况下

6 关于证明参见 Arrow（1962）。

的创新激励，但是即使在后一种情形下，创新的激励仍然低于社会最优情况”（Arrow，1962，p. 152）。在阿罗的断言中至少有两个不足之处。首先，分析只限制在产品市场的供给方的不同情形，而没有考虑在其他环境下进行研发活动的情况。如果后者被视为来自创新活动的净收益，那么这可能不能满足不等式（2）。特别地，在关于竞争性情形的模型化中，似乎他假设在研发活动中是不存在竞争的，并且只有一家企业从事研发工作（见脚注6）。其次，该模型只假设不连续选择是可行的。更一般地，人们会假设存在边际选择：研究战略的目的在于生产成本的少许降低，风险的少许增加，速度的少许加快。在接下来的三个部分中，我们将依次考虑过程创新的这三种特征。

二、市场结构与研发支出的大小

1. 社会经营产业

首先，考虑我们所研究的商品市场是社会经营（socially managed）的情形。如果单位生产成本为 c（常数），那么消费 Q 的净社会收益为 $u(Q)-cQ$。定义

$$V(c) = \max_{Q}[u(Q) - cQ] \tag{3}$$

为间接效用函数。我们都知道这是一个关于 c 递减的凸函数（见图4）。我们假设研发支出的目的在于降低生产成本。如果 x 为研发支出水平，那么 $c(x)$ 就是生产的单位成本。在这里，很自然地我们假设 $c'(x)<0$。[7]

假设政府可以通过一般的征税方式募集研发所需的资金。那么政府所面临的问题为

$$\max_{x\geqslant 0}\{V[c(x)] - x\} \tag{4}$$

假设式（4）存在一个解。我们需要注意的第一点就是：除非存在关于 $c(x)$ 性质的进一步限制，不然我们不能保证 $V[c(x)]$ 为关于 x 的一个凹函数。即使当我们假设 $c''(x)>0$（技术研发的回报递减）时，仍然是这样。函数 V 可能为非凹函数的问题是一个很常见的问题。假设式（4）的解表明 $x>0$，那么最优的研发支出水平必须满足

$$\frac{dV}{dc}c'(x) = 1$$

7 就好像上帝拥有具有单位成本为 $c(x)$（$x>0$）的所有生产技术的专利，而企业必须支付 x 元以获得该具有单位生产成本为 $c(x)$ 的生产技术的使用权。

但是

$$\frac{dV}{dc} = -Q$$

从而

$$-c'(x)Q = 1 \qquad (5)$$

式（5）的含义是很清楚的。左边是增加研发支出的边际社会收益，事实上，这就是当产出水平最优时生产成本的节省。显然，方程右边就是增加研发支出的边际成本。如果函数 V 为关于 x 的凹函数，那么政府要确定最优并没有必要计算消费者剩余。但是当 V 不是凹函数时，一般而言，政府都需要进行全局的成本—收益分析来确定最优解的位置。在这种情况下，计算消费者剩余的必要性是显然的。

为了更清楚地说明这些要点，对函数形式做出一定的具体化被证明是有帮助的。假设

$$u(Q) = \frac{\sigma Q^{1-\epsilon}}{(1-\epsilon)}(\sigma,\epsilon > 0) \qquad (6)$$

所以，如果我们记市场需求曲线 $p(Q)$ 为

$$p(Q) = \sigma Q^{-\epsilon} \qquad (7)$$

其中 ϵ^{-1} 为需求弹性。所以，实际上式（7）相当于假设市场需求具有不变弹性。

此时，很容易地，可以证明

$$V[c(x)] = \left(\frac{\epsilon}{1-\epsilon}\right)(\sigma)^{1/\epsilon}/[c(x)]^{(1-\epsilon)/\epsilon} \qquad (8)$$

另外，假设我们所考虑的商品是一种新产品。特别地，我们假设

$$c(x) = \beta x^{-\alpha}(\alpha,\beta > 0) \qquad (9)$$

在这种情况下，

$$V[c(x)] = \left(\frac{\epsilon}{1-\epsilon}\right)\frac{\sigma^{1/\epsilon}}{\beta^{(1-\epsilon)/\epsilon}}x^{\alpha(1-\epsilon)/\epsilon}$$

现在，我们可以看到当 $\epsilon > \alpha(1-\epsilon)$ 时函数 $V[c(x)]$ 为关于 x 的严格凹函数。如果该不等式的反方向成立，$V[c(x)]$ 为关于 x 的严格凸函数，而且，特别地，式（4）将不存在有限解。社会收益会随着研发支出的增加而一直无限增加。[8]

8 显然，这是很荒唐的，因为如果我们假设收入的边际效用保持为一个常数是很不合理的。

我们现在得到的关于保证 $V[c(x)]$ 为 x 的严格凹函数的条件含义是相当清楚的。它意味着 $1+1/\alpha>1/\epsilon$。因为 $1/\epsilon$ 为需求弹性，而且 $1/\alpha$ 为研发投资的生产成本弹性，那么这个条件说明，后者这个弹性必须要足够的大，以使得研发支出的（效用）回报连续递减。应当注意到当需求是缺乏弹性时（即 $\epsilon>1$），该条件确实能够被满足。而且即使需求是富有弹性时（$\epsilon<1$），该条件也是有可能被满足的，当然只要其弹性不是很大。在收入弹性为零的条件下，这个条件能够使一个关于收入效应为零假设、良好定义的计划问题 ϵ 的取值区间精确化。[9]

现在假设 $\epsilon>\alpha(1-\epsilon)$。令 x_s 和 Q_s 分别表示最优的研发支出水平和最优的产出水平。如果我们现在利用最优条件中的式（7）和式（9）很容易证明

$$x_s=(\alpha^{\epsilon}\sigma\beta^{\epsilon-1})^{1/[\epsilon-\alpha(1-\epsilon)]} \tag{10}$$

和

$$Q_s=(\alpha\beta)^{-1}(\sigma\alpha^{\epsilon}\beta^{\epsilon-1})^{(1+\alpha)/[\epsilon-\alpha(1-\epsilon)]} \tag{11}$$

根据式（10）和式（11）马上可以得到，拥有较大规模的市场（也就是说，σ 值较大）的行业的最优产量和最优的研发支出水平都是较高的。所以市场规模越大（单位）成本降低应该越多。同样值得注意的一点是，在那些研发技术的成本较高（即具有较高的 β 值）的行业中总产出水平应该较低。也许有人预期这些就是最优的特征。但是，式（10）意味着如果需求是缺乏弹性的（$\epsilon<1$），那么研发技术成本较高的行业的最优研发支出将会较低，而如果需求是富有弹性的（$\epsilon>1$）那么其最优研发支出就会较高。仔细考虑一下，我们会发现，这也是与我们的直觉相吻合的。

2. 自由进入条件下的寡头竞争

仍然考虑一个市场经济。现在假定研发活动只在私人部门进行。为了分析的简单，我们考虑一种新产品的市场，同样是出于简单的考虑，我们假设

9　到目前为止，我们分析的目的在于将注意力集中在某些环境当中，过程创新活动的收益函数可能的非凸性存在的问题；不再有其他的。就像我们在前面的脚注 8 中所提到的那样，在假设不存在收入效应的同时考虑非凸性的做法是相当不明智的。最清楚的问题在于当事实上 $\epsilon<\alpha(1-\epsilon)$ 时，如何重新对我们的模型进行描述。我们显然需要假设收入的边际效用不再是常数。例如，用一种很清楚的方式，我们假设 $u(Q,x)=\sigma[Q^{(1-\epsilon)}/(1-\epsilon)][M-x-c(x)Q]^{\delta}$，其中 $1>\epsilon$，$\delta>0$ 而 M 为总收入。令 $W(x)=\max_{Q}u(Q,x)$ 我们可以很简单地证明，如果 $1<\alpha(1-\epsilon)$，当 $c(x)$ 满足式（9）时，那么 $W(x)$ 关于 x 的小量为严格凸函数。但是，$W(x)$ 具有一个唯一的最大值。

所有的企业都面临同样的研发技术 $c(x)$。所以，如果 $x_i(\geqslant 0)$ 为企业 i 的研发支出水平，那么 $c(x_i)$ 就是生产时所面临的单位生产成本。在这里我们假设，如果一家公司对某知识进行了支付，那么它就拥有了对于该知识的垄断权。[10]企业追求利润最大化，非合作地各自采取行动，而且我们的任务就是为这样一个市场定义均衡。为了实现这一目的，我们假设每一家企业都以一种古诺的方式采取行动（每一家企业决定自己的研发支出水平和产出水平），而且所有的企业都接受古诺猜想（所有的公司都认为，如果给定的某家企业改变自己的行为，其他的企业都不会偏离它们原来的行动）。

我们首先假设该行业允许自由进入，也就是说，假设市场结构是内生的。令 n 表示该行业中的企业个数而 Q_i（$\geqslant 0$）表示第 i 家企业的产出水平（$i=1$，…，n）。根据我们的假设有

$$[n^*, (Q_1^*, x_1^*), (Q_2^*, x_2^*), \cdots, (Q_i^*, x_i^*), \cdots, (Q_{n^*}^*, x_{n^*}^*)]$$

为自由进入条件下的一个均衡，如果，对于 $i=1$，…，n^* 有

$$[p(\sum_{j\neq i} Q_j^* + Q_i^*) - c(x_i^*)]Q_i^* - x_i^* \geqslant [p(\sum_{j\neq i} Q_j^* + Q) - c(x_i)]Q_i - x_i$$

$$\text{任意 } x_i, Q_i \geqslant 0 \tag{12}$$

而且

$$[p(\sum_{i=1}^{n^*} Q_i^* + Q) - c(x)]Q - x \leqslant 0, \text{任意 } x, Q \geqslant 0 \tag{13}$$

条件（12）的含义是很清楚的。它意味着此时的决策（Q_i^*，x_i^*）就是企业 i 的利润最大化选择，如果企业 i 认为处于该行业中的其他的企业 $j(j\neq i)$，都选择产量水平 Q_j^*，且其他不处于该行业中的企业都不进入该行业。如果该行业不存在进入的障碍，那么条件（13）必须得到满足。它意味着如果该行业中的企业 i（$i=1$，…，n^*）选择产量水平 Q_i^*，那么，一家原本不在该行业中的企业是无法进入该行业而获得正的利润的。所以，这里我们其实假设了所有的企业，不管是在该行业中，还是不在该行业中，在考虑所有其他企业的行动时都接受古诺假设。

这里，我们不但关注一个自由进入的市场均衡是否存在的问题，而且还关注在那些存在进入障碍的情况下均衡具有什么特征。这后一个任务将会非

10　由于该产品为新产品，那么根据假设有 $ci(0)=\infty$。关于当现存的竞争性产品价格为 c，而且企业花费资源去降低成本（也就是说，$c_i(0)=c$）的情形我们留给读者去分析。关于这个模型的另外一种可选的理解方式为，假设偏好是定义在商品的特征上的，而且该行业中的企业通过生产不同的，但是具有同样特征的商品以竞争。所以这些商品就是消费方面的完全替代。

常简单，如果我们只将分析限制在对称均衡的范围内，也就是，均衡时行业中所有的公司都采取同样的行动。在附录 I 中我们会给出在哪些情况下对称均衡可以被证明是存在的。现在，我们假设这些均衡是存在的，从而，我们可以进一步考察其特征。

令 n^* 为均衡时行业中的企业数目（当然是需要确定的）。令 $\hat{Q}_i$ 表示除企业 i 之外所有其他企业的总产出水平。从而 $Q = Q_i + \hat{Q}_i$。企业 i 选择 x_i 和 Q_i 使下式最大化：

$$\max\{[p(Q_i + \hat{Q}_i) - c(x_i)]Q_i - x_i\} \tag{14}$$

根据利润最大化的 x_i 和 Q_i 都为正的假设，我们有一阶条件

$$p(Q)[1 - \epsilon(Q)Q_i/Q] = c(x_i) \tag{15}$$

和

$$-Q_i c'(x_i) = 1 \tag{16}$$

其中 $\varepsilon(Q) \equiv -Qp'(Q)/p(Q)$，为需求弹性的倒数。既然我们这里考虑的是对称均衡，那么 Q_i 和 x_i 就应当独立于 i 对于 $i = 1, \cdots, n^*$。所以，如果 n^*，Q^*，x^* 刻画了一个自由进入条件下的市场均衡，那么它们必须满足条件式（15）和式（16），此时条件式（15）和式（16）变成了

$$p(Q^*)[1 - \epsilon(Q^*)/n^*] = c(x^*) \tag{17}$$

和

$$-c'(x^*)Q^*/n^* = 1 \tag{18}$$

（注意到 Q^* 表示均衡时的行业总产出水平，那么 Q^*/n^* 就是此时行业中每一家企业的产出水平。）

更进一步地，自由进入的条件式（13）变为

$$[p(Q^* + Q) - c(x)]Q - x \leqslant 0, \text{对于所有的} x, Q \geqslant 0 \tag{19}$$

最后注意到（x^*，Q^*/n^*）为均衡时行业中具有代表性企业的利润最大化选择，那么它必须产生非负的利润，即

$$[p(Q^*) - c(x^*)]Q^* \geqslant n^* x^* \tag{20}$$

显然，如果自由进入的条件能够导出每一家企业都赚取零利润，那么分析就能被大大地简化。现在假设 $\{[p(Q^*) - c(x^*)]Q^* - n^* x^*\}/n^* x^*$ 小到可以被忽略。从而我们可以不再专注于式（20），而是利用零利润条件：

$$[p(Q^*) - c(x^*)]Q^* = n^* x^* \tag{21}$$

如果条件（21）能够得到满足，那么式（19）显然也是满足的。但是我们也

可以得到即使当不等式（20）的不等号严格成立时，式（19）也能得到满足的条件。我们将在附录I中详细讨论这个问题，在那里我们还将讨论在什么条件下式（21）是一个良好的近似。事实上有人可能会认为当 n^* 较大时式（21）是一个良好的近似，而且也可以确认均衡时企业数目较大的参数条件。在附录I中，我们将通过一类非常重要的情形来说明事实确实是这样的，但是在有些将式（21）作为均衡条件是很合理的条件下 n^* 并不一定很大。这里我们就假设式（21）是一个良好的近似。从而我们可以从式（17）、式（18）和式（21）中解出 n^*，Q^*，x^*。

初步审视一下均衡条件，我们发现市场均衡并没有达到资源的最优配置。首先，在一个保持一定生产者们研发投资水平的市场均衡中，生产者具有某种程度的垄断力量。其次，私人企业的研发支出的边际收益是基于其对该企业本身的产出规模的贡献（即条件式（18））而计算的，而不是基于其对整个市场的贡献（条件式（5））来计算的。

接下来，我们详细分析市场均衡的条件。在利用式（17）和式（21）时我们注意到

$$1/n^* = Z^*/\epsilon(Q^*)(\text{其中 } Z^* = n^*x^*/p(Q^*)Q^*) \tag{22}$$

式（22）是这一部分的基本等式。因为我们现在分析的是对称均衡，所以我们不能使用集中比例等参数来比较不同市场的集中程度。对于我们的模型，很自然地，我们就用 $1/n^*$ 作为市场集中指数。Z^* 为使用在研发上的行业产出规模的比例，是该行业研究密度的一个很明显的指数。式（22）说的是如果 ϵ 为一个常数，那么它们之间互成比例。所以，在对于具有相同的均衡需求弹性，但是，在具有不同的市场规模和研发技术的行业的研究中，我们将能看到研究密度与市场集中度之间具有线性关系。但是这种关系并不具有什么因果关系的成分：产业集中度与研究密度是同时决定的。[11]

注意到在得到式（22）的过程中，我们并没有用到式（18）。所以我们

11 这里得到的行业集中度与研发支出之间的关系与脚注3中提到的实证结果并不矛盾，因为注意到式（22）是在式（21）是一个良好近似的假设下才得到的。在附录I中我们将要看到对于模型中参数大小的一定的区间和，特别是那些使得 n^* 较大（市场集中度较低）的变化区间，式（21）都是一个良好的近似。为了让大家对与大小顺序有一个直观的感受，我们可以注意到对于1961年美国的公司样本中，那些雇佣超过5 000名员工的大公司，平均的 Z 值为5.2%，而对于那些雇用员工在1 000～5 000名之间的公司为2.2%，对于那些雇佣员工低于1 000人的小公司为2%（见Nelson等人，1967，p.67）。

式（22）能够在比允许条件的更一般的条件下得到，它并不依赖于公司根据利润最大化的考虑选择研发投资水平的假设。所以我们接下来将尝试导出一个更尖锐的特点。令 $\alpha(x) \equiv -xc'(x)/c(x)$ 表示生产单位成本关于研发支出的弹性。然后根据式（18）、式（22）有

$$Z^* \equiv n^* x^* / p(Q^*) Q^* = \alpha(x^*) / [1 + \alpha(x^*)] \tag{23}$$

所以，在一个对具有不同均衡需求弹性，但是具有相同单位成本弹性函数（α）的不同产业的研究中，我们可以看到，各产业的产业研究密度 Z^* 是相同的。但是这些产业的产业集中度可能是不同的；因为利用式（23）、式（22）我们有，均衡时的企业数目为

$$n^* = \epsilon(Q^*)[1 + \alpha(x^*)] / \alpha(x^*) \tag{24}$$

因此，在这样的跨部门研究中，部门的需求弹性（$1/\epsilon(Q^*)$）越大，则我们观察到的均衡企业数目就越小。

但是，注意到式（22）和式（24）告诉我们，在任意给定的公司数目的条件下，具有较小需求弹性的行业将会有较高的行业研究密度指数。在给定的市场集中度条件下，较低的需求弹性将导致较高的价格；均衡并不是通过进入来维持的，而是通过企业在研发上支出足够高来阻止进入来维持的（事实上是通过将它们所有的利润都投在研发上）。

目前，这些命题都具有明显的熊彼特风格。既然研发是有固定成本的，那么我们就不能期望一个从事研发的行业会具有完全竞争的特征。但是，通过企业不停地进入，该行业可以保持有效竞争。我们模型中的诸如市场价格超过生产单位成本（式 17）之类的限制条件必须在熊彼特称之为"创造性破坏之风"（perennial gale of creative destruction）的环境中加以理解。我们的分析显示，一个行业中的企业个数并不能度量这种有效竞争的程度。同时，市场规模也不能直接影响行业中的企业数目，因为正如式（24）所示，均衡企业数目只受到需求弹性和创新函数的影响。如果如式（7）、式（9）所示那样，这些弹性都是不变常数，那么均衡企业数就可以直接算出，而且与市场规模无关。但是，市场规模的大小确实影响在一个市场经济中过程创新发生的程度。为了更明显地看到这一点，我们假设市场需求与创新函数满足式（7）和式（9）。如果现在我们在均衡条件式（17）、式（18）和式（21）中利用这些函数形式，那么常规的计算得到，其解为

$$n^* = \epsilon(1 + \alpha) / \alpha \tag{25}$$

$$Q^* = \frac{\epsilon(1+\alpha)}{\alpha^2\beta}[\sigma\alpha^{2\epsilon}\beta^{\epsilon-1}\epsilon^{-\epsilon}(1+\alpha)^{-(1+\epsilon)}]^{(1+\alpha)/[\epsilon-\alpha(1-\epsilon)]} \quad (26)$$

和

$$x^* = [\sigma\alpha^{2\epsilon}\beta^{\epsilon-1}\epsilon^{-\epsilon}(1+\alpha)^{-(1+\epsilon)}]^{1/[\epsilon-\alpha(1-\epsilon)]} \quad (27)$$

现在，回顾一下条件（17）、条件（18）仅仅是关于每一家公司利润最大化实践的一阶条件。在附录Ⅰ中我们将看到，要使条件（25）到条件（27）定义一个均衡，我们需要假设$\epsilon>\alpha(1-\epsilon)$（如我们前面所提到的那样，当市场需求缺乏弹性时（$\epsilon<1$）这一点是自然成立的）。然后根据式（27）立即有，市场的规模越大（σ越大），那么平均每一家公司的研发支出就越大，从而单位成本的降低也就越大。从而均衡的行业总产量也就越大。[12]类似地，研发技术的成本越高（β越大），那么行业总产出也就越小。这也是与我们的直觉相吻合的。但是应注意到，如果需求是富有弹性的，那么研发技术成本越高，则均衡时每一家公司的研发支出也就越小（同样总的行业研发支出也就越小）；而当需求缺乏弹性时，研发支出却越高，这并不是一个直觉上很明显的结论。

注意到我们得到式（25）至式（27）是在假设均衡中的企业赚取零利润的前提条件下推导得到的。那么显然当$\epsilon(1+\alpha)/\alpha$为一个整数时，将存在一个自由进入的对称均衡，而且此时均衡中的企业赚取零利润，而均衡时行业中的企业个数由式（25）决定。[13]此时，虽然企业数目必须为整数，但是一般来说，$\epsilon(1+\alpha)/\alpha$不会是整数。但是，如果$\epsilon/\alpha$"较大"，那么不超过$\epsilon(1+\alpha)/\alpha$的最大整数也较大。此时条件（25）意味着如果我们令$n^* = [\epsilon(1+\alpha)/\alpha]$（即不超过$\epsilon(1+\alpha)/\alpha$的最大整数），那么这么多数目的企业可以维持一个均衡，因为此时如果每一家企业选择满足条件（26）、条件（27）的$(x^*, Q^*/n^*)$作为决策，那么给定其他企业的决策，每一家企业都能够使利润水平最大化，而且最大化的利润水平如此之低使得条件（19）能够被满足，从而不会有进一步的进入。在附录Ⅰ中，我们将看到事实上确实是这样的。根据分析我们知道，如果ϵ/α"较大"，那么一个自由进入的对称均衡是存在的，而且通过选择要么是充分小的α，要么是充分大的ϵ，可以使得每一家企业的均衡利润水平达到我们想要的低，从而使得零

12 后来的雅各布·施穆克勒（Jacob Schmookler）在他的一系列作品中强调了在刺激以降低生产成本和提高产品质量为目的的研发支出中，对一种产品的需求的增长的重要性。例如，见 Jacob Schmookler（1962）。

13 由于我们已经假设了$\epsilon>\alpha(1-\epsilon)$，从而有$\epsilon(1+\alpha)/\alpha>1$。

利润条件（21）如我们所愿成为一个良好近似。[14]但是应当注意到 α 和 ϵ 的两个极限值将导致行业均衡的不同特点。为了说明这一点，在式（17）中利用式（24），得到

$$p(Q^*)/c(x^*) = 1 + \alpha(x^*) \tag{28}$$

假设 α 和 ϵ 都是常数。注意到当 $\alpha \to 0$ 时 $\epsilon/n^* \to 0$，而且根据式（28）我们得到，当 $\alpha \to 0$ 时，$p(Q^*) \to c(x^*) = \beta$。当 $\alpha \to 0$ 时，在极限处这个模型就退化成为一个传统的完全竞争行业的模型。但是，当 $\epsilon \to \infty$ 时，$\epsilon/n^* \to \alpha(1+\alpha)$，而且特别地，式（28）意味着，如果行业需求高度缺乏弹性，那么市场均衡中将存在很多家公司。但是此时市场价格与边际生产成本之比可能比 1 大很多。

在前面我们提到，对于目前我们这个模型，使用 n^{-1} 作为市场集中度的指数是很自然的做法。事实上，直觉告诉我们，如果均衡时行业中企业数目较大，那么这个行业就类似传统的竞争性行业。但是现在的模型结构告诉我们，这种直觉是错误的。实际上，曾经有作者（例如，Kalecki，1954）使用生产价格对“原始成本”（在这里就是 $p(Q)/c(x)$）之比作为一个行业的垄断程度的度量。现在，通过式（28）我们可以发现一个行业可以同时具有低的集中度（较大的 n^*）和高的垄断程度（较大的 $p(Q^*)/c(x^*)$）的特征。当 ϵ 和 α 都“较大”，从而 $n^* \equiv [\epsilon(1+\alpha)/\alpha] \simeq \epsilon$，从而 n^*“较大”，此时就会出现这种情况。事实上，对于我们这里的模型，我们将看到垄断程度是比集中程度关于市场不完美性更好的度量。

现在，在对一些具有不同的单位生产成本弹性的行业的分析中，我们会看到那些具有较大弹性的行业具有较高的垄断特征。如果这些行业面对着同样的需求弹性，那么式（17）告诉我们，具有较高垄断程度的行业将会有较少的均衡行业企业数目。但是，通过式（22）我们知道，这又是与高的行业研究密度 Z^* 相联系的。那么现在的问题是到底每一家公司的研发支出是较高还是较低？一般而言，这是很难做出明确判断的。但是，如果假设 α 和 ϵ 都为常数，从而均衡可以由式（25）至式（27）表述。现在应当注意到，根据式（27），如果 α 较小且需求缺乏弹性（$\epsilon > 1$），那么 $\partial x^*/\partial\alpha > 0$，从而，此时 α 越大，则寡头均衡时每家公司的研发支出也就越高。但这同

14　要使式（17）有意义 n^* 必须大于 $\epsilon(Q^*)$。但是，如果 ϵ 和 α 都是常数，那么当或者 α“较小”或者 ϵ“较大”时 $n^* = [\epsilon(1+\alpha)/\alpha] > \epsilon$。这里以及之后的部分，一个数字外面加黑方括号表示不超过该数字的最大整数。

时又意味着该行业的单位成本降低幅度较大。那么，让人感到有点自相矛盾的是，我们将在那些具有较高垄断程度的行业中观察到较大生产成本的降低。

我们需要比较的是，自由进入的寡头竞争均衡与社会经营行业的均衡。我们仍然假设 $p(Q)$ 和 $c(x)$ 分别满足式（7）和式（9）。然后，我们注意到根据式（11）和式（26）有：$Q_s > Q^*$。市场均衡产出水平要小于社会最优的产出水平，这个结果从直觉上来看并不很明显，因为人们往往会认为，竞争（自由进入）的压力会通过迫使各厂商在研发方面的支出超过社会最优水平，从而使得市场价格在一定程度上低于社会最优水平。于是，我们想问，寡头市场上的生产成本降低是高于还是低于社会最优。注意到，$(n^*)^{-\epsilon} \gtrless (1+\alpha)$ 时 $x^* \gtrless x_s$，比较式（10）和式（27）。我们得到 $x^* < x_s$，从而也就是说，在寡头竞争市场上的生产成本降低程度不够高。但是，在该市场经济中的行业范围内的研发投资水平为 $n^* x^*$，我们的问题就是如何比较这个值与 x_s。现在根据式（10）、式（25）和式（27），我们可以较容易地得到：当 $(n^*)^{\alpha(\epsilon-1)} \gtrless (1+\alpha)$ 时，$n^* x^* \gtrless x_s$。这意味着，如果 ϵ “较大”（即需求是高度的缺乏弹性），那么 $n^* x^* > x_s$，从而整个行业范围内的研发投资水平超过社会最优水平。因此，与社会经营的经济相比，市场经济可能同时具有过度的研发支出（$n^* x^*$）和过低的技术进步率（x^*）这两大特征。这里的关键当然在于市场鼓励过度的重复劳动。为了更明显地看到这一点，我们注意到根据式（10），当 $\epsilon \to \infty$ 时，$x_s \to (\alpha\beta)^{1/(1+\alpha)}$。而从式（27）直接可得当 $\epsilon \to \infty$ 时 $x^* \to 0$，但是同时 $n^* x^* \to \infty$。所以当 $\epsilon \to \infty$ 时，由竞争导致的福利损失趋向于无穷，即使当 $\epsilon \to \infty$ 每一家公司只面对有限需求的一个无穷小的比例。

而在另一个极端情况就大不一样了。根据式（10）和式（27）立即可得，当 $\alpha \to 0$ 时 $x^* \to 0$ 且 $x_s \to 0$。同样也注意到，当 $\alpha \to 0$ 时 $n^* x^* \to 0$，即使 $n^* \to \infty$。所以，当 $\alpha \to 0$ 时，在极限处由于该行业私营化所导致的福利损失趋于零。正如我们前面所说的，此时该行业就是传统的竞争性行业，不存在扭曲。此时竞争性的福利损失为零。

3. 存在进入障碍的寡头竞争

本文的研究目的就是，考察一个内生的市场结构对于创新活动的数量的关系。在模型中，我们通过自由进入条件（13）刻画了这个关系。进入保持了有效的竞争。正如我们在导言部分中提到的那样，这意味着创新活动的节奏既不能通过行业集中度来理解，也不能通过行业的垄断程度来

理解，而应该通过一些更基本的因素，诸如需求条件、研发技术、资本市场本质等来研究。然而，有些作者却宣称在发达的资本主义经济中创新的密度是逐渐下降的，而且他们认为这是与资本主义经济逐渐加强的垄断特征相联系的。[15]

现在，我们已经在很多场合说明了，行业的垄断程度 $p(Q)/c(x)$ 本身并不是一个解释变量。同样，如果存在自由进入，行业中的企业数目也不能作为一个解释变量。事实上，我们可以看到，在可以自由进入该行业的条件下，那些具有较高程度垄断的行业完全有可能更具有创新性，而不是更缺乏创新性。因此现在我们就很有必要在存在进入障碍的环境中，深入探讨一下垄断程度与生产成本的降低程度之间的关系。然后，我们就可以考察允许新企业的进入是否会导致均衡时更高的成本降低程度。我们先令企业数目为外生给定的，比如 n。正式地，如果对于 $i=1, \cdots, n$

$$\left.\begin{aligned}&[p(\sum_{j\neq i} Q_j^* + Q_i^*) - c(x_i^*)]Q_i^* - x_i^*\\ &\geqslant [p(\sum_{j\neq i} Q_j^* + Q_i) - c(x_i)]Q_i - x_i, \text{对于所有 } x_i, Q_i \geqslant 0\end{aligned}\right\} \quad (29)$$

那么 $[(Q_1^*, x_1^*), \ldots, (Q_i^*, x_i^*), \ldots, (Q_n^*, x_n^*)]$ 为一个均衡。既然我们只对对称的均衡感兴趣，那么当前的均衡的定义意味着条件（17）和条件（18）必须得到满足，这里，我们将其重写为

$$p(Q^*)[1 - \epsilon(Q^*)/n] = c(x^*) \quad (30)$$

和

$$-c'(x^*)Q^*/n = 1 \quad (31)$$

因为 n 是给定的，那么式（30）和式（31）中只有两个未知数，Q^* 和 x^* 需要确定。为了得到显式解，我们假设 $p(Q)$ 和 $c(x)$ 分别满足式（7）和式（9）。根据式（30），显然，因为现在我们假定 n 是外生给定的，所以我们必须假设 $n>\epsilon$（如果不然，那么式（30）和式（31）将无解）。同样，我们还必须假设 $n\leqslant\epsilon(1+\alpha)/\alpha$，因为如果不然，式（30）和式（31）的解将导致公司的利润为负，从而不能满足均衡条件。最后，我们还需要假设 $\epsilon>\alpha(1-\epsilon)$，该条件在前面的分析中也是必要的（详见附录I）。然后通过常规计算我们可以解得式（30）和式（31）的解为

$$x^* = [\sigma(\alpha/n)^{\epsilon}\beta^{\epsilon-1}(1-\epsilon/n)]^{1/[\epsilon-\alpha(1-\epsilon)]} \quad (32)$$

15　“另一个（导致发达的资本主义经济增长过慢的原因）就是资本主义经济逐渐加强的垄断特征阻碍了新发明在实际中的应用”（Kalecki，1954，p. 159）。

和

$$Q^* = (n/\alpha\beta)[\sigma(\alpha/n)^{\epsilon}\beta^{\epsilon-1}(1-\epsilon/n)]^{(1+\alpha)/[\epsilon-\alpha(1-\epsilon)]} \quad (33)$$

从式（33）我们马上可以看到在 n 的可接受变化范围（ϵ，$\epsilon(1+\alpha)/\alpha$）内，$Q^*$ 为关于 n 的增函数。因而，行业产出水平随着行业中企业数目的增加而增加，从而产品价格下降。这正是竞争程度更高的优点所在。根据式(30)，很明显的是：当行业中企业数目增加时，行业的垄断程度降低。而我们开始想要搞清楚的问题为：是否企业数目的增加会导致更多的创新活动(即更大程度的成本降低)。由式（32）答案显然是“否”。如果企业数目增加，那么均衡时每一家企业花费在研发方面的支出就更少，均衡时的单位生产成本就更高些。但是，由式（32）也很容易验证总的研发支出 nx^*，会随着企业数目的增加而增加。关键在于虽然就整个行业而言，由于更激烈的竞争在研发方面的支出更多了，但是每一家企业却花得更少了。许多额外的支出都浪费在了重复研究上面。事实上，如果我们对照式（10）和式（32）就可以得到 $x^* < x_s$。因此对于我们这个模型有如下结论，不管是否存在进入障碍，市场经济中的（单位）成本降低都是无效率的，从而导致产品的市场价格较社会最优水平要高。

如果考虑一种受到进入门槛保护的垄断者情形，那么我们能很清楚地看到这一点。在进入门槛保护的情形下，当然不再存在研发方面的重复支出。但是，我们已经看到了此时垄断者从事的研发活动要低于社会最优水平。即使在一个社会经营的行业中，政府被迫通过征收利润税来提高研发方面的支出，这个问题仍然存在。

然而，如果可以自由进入研发活动时，公司展开研发工作的速度，以及之后所导致的市场经济中技术更新的速度，将要比当研究部门对外来者存在进入门槛时要快些。在第四部分我们将考虑研究部门自由进入条件对于研发活动进行速度的影响。关于该问题的一个更完整的讨论请参见达斯古普塔和斯蒂格利茨（1980）的论文。

三、过程创新与风险承担度的非凹性

我们已经注意到式（3）中的社会间接效用函数 $V(c)$ 为递减的凸函数(见图4)。我们回到第一部分的基本模型，然后假设 c 为实践中最好生产技术的单位生产成本。我们的目的就是要考虑函数 $V(c)$ 的凸性对于具有风险性的研究项目的选择的影响。首先，我们假设有两种可供选择的研究战略，其中一种一定可以把生产成本降至 c^*，而另外一种，如果成功，则能够将

生产成本降至 $\tilde{c}(\tilde{c}<c^*<c)$。而如果不成功，则生产成本仍然维持为 c 的水平。但是假设风险研究项目的期望成本降低幅度为 $c-c^*$。如果社会福利标准为最大化社会期望效用水平，那么，如果风险项目与无风险项目成本是一样，则风险项目将优于无风险项目（见图 4）。[16] 现在我们来将该例子一般化。

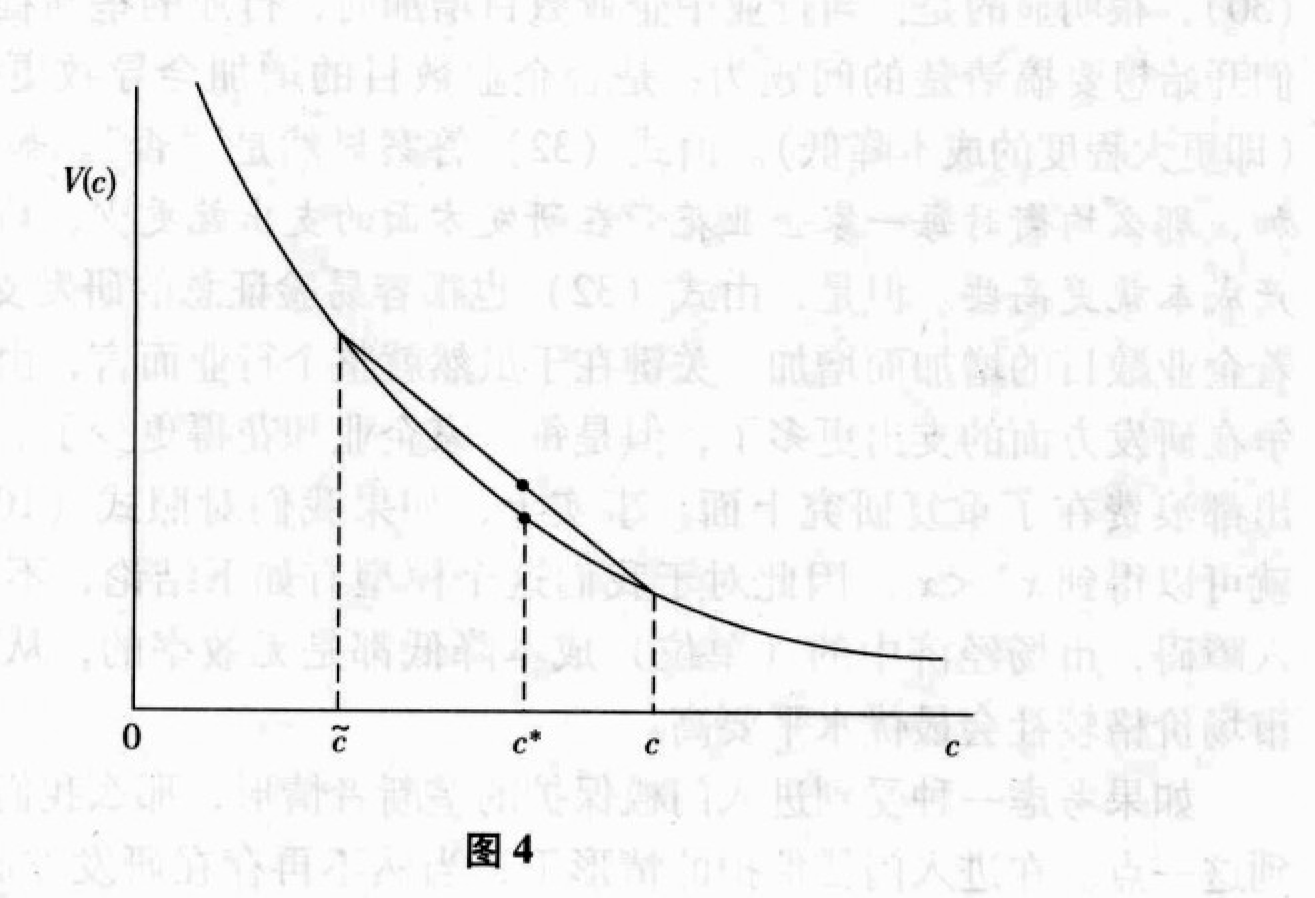


图 4

假设研发项目由 α 表示（$\bar{\alpha}\geqslant\alpha\geqslant0$）。项目 α 具有的成功概率为 $h(\alpha)$。如果成功那么单位生产成本将降至 $c(\alpha)$。如果不成功那么单位生产成本保持为 c。为了不失一般性，我们假设 $c'(\alpha)<0$。出于简单的考虑，我们仍然假设这一族研发项目都具有相同的期望产出；即

$$\left.\begin{aligned}&E[c(\alpha)]=c^*=h(\alpha)c(\alpha)+[1-h(\alpha)c],\\&\bar{\alpha}\geqslant\alpha\geqslant0\quad 且\quad h(0)=1\end{aligned}\right\}\tag{34}$$

式（34）意味着 $c(0)=c^*$，从而 $\alpha=0$ 就是无风险项目。如果我们对式（34）求一阶导数，得到

$$c'(\alpha)h(\alpha)=[c-c(\alpha)]h'(\alpha)\tag{35}$$

从而 $h'(\alpha)<0$。所以 α 值越大意味着该项目的风险就越高。

项目 α 的研发支出为 $R(\alpha)$。显然 $R(\alpha)>0$，$\alpha\geqslant0$。如果 $R'(\alpha)\leqslant0$，那么就不再有讨论的意义。既然 $\bar{\alpha}$ 是可获得的风险最大的项目（在给定这

16　如果研发成本过高，那么社会将不会愿意进行任何研发工作，而只是依赖于原来的 c—生产技术。

一族研发项目中最大的 α 值）那么计划者应该选择 $\bar{\alpha}$（或者根本不进行任何形式的研发项目）。我们考虑另外一种 $R'(\alpha)>0$ 的特殊的情形；为了明白起见，我们再次假设 $R''(\alpha)>0$。

在接下来的部分，我们假设 $R(0)$“较小”，从而最优情况也包含一些研发活动。现在，如果计划者选择项目 α，那么净期望社会收益 $E(\alpha)$，为

$$E(\alpha)\equiv h(\alpha)V[c(\alpha)]+[1-h(\alpha)]V(c)-R(\alpha) \qquad (36)$$

注意到，除非我们对于函数 $h(g)$ 和 $V(g)$ 的结构形式做出进一步的限定，我们不能保证函数 $E(g)$ 一定是凹的。如果不是那样的话，那么一般来说就需要政府“随机”对研发项目的全局进行成本—收益分析。

问题已经提出，现在我们来做一些简化，假设式（36）是严格凹的。假设最优的 α 值严格地落在 $\bar{\alpha}$ 和零之间，然后根据式（35），我们知道最优的充分必要条件为

$$-h(\alpha)c'(\alpha)\left(-V'[c(\alpha)]-\frac{\{V[c(\alpha)]-V(c)\}}{[c-c(\alpha)]}\right)=R'(\alpha) \qquad (37)$$

社会成本—收益标准式（37）的含义异常地简单明了。$-h(\alpha)\ c'(\alpha)$ 为当一个边际上更具风险的研发项目被选择时，期望的边际生产成本下降。式（37）左边部分的含义就是提高研发支出的边际收益为 $-h(\alpha)\ c'(\alpha)$ 乘以成本降低的边际社会收益 $\{-V'[c(\alpha)]\}$，与平均降低的平均社会收益 $\{V[c(\alpha)]-V(c)\}/[c-c(\alpha)]$ 之差。

下面从另外一种方式看这个问题，我们知道

$$-V'[c(\alpha)]>\{V[c(\alpha)]-V(c)\}/[c-c(\alpha)]$$

（因为 $V(c)$ 递减且凸）。那么式(37）的含义就是，最优时边际的和平均的回报之差应当为 $-R'(\alpha)\ /h(\alpha)\ c'(\alpha)$。

我们已经得到了一个社会中央计划者愿意从事更具风险的研发项目的结论。但是一个纯粹的垄断者也会这样做的。为了说明这一点，我们假设销售的边际收入为关于销售额的一个减函数。那么，我们就可以很容易证明，垄断者的最大化利润为单位生产成本的凸且递减的函数。因此，如果垄断者关注于最大化期望利润水平，用与前面我们使用过的分析方法完全类似的方法，我们可以得到垄断者也将偏好于风险。

很自然地，我们想要问的一个问题是，垄断者是否倾向于承担最优的风险程度，如果不是的话，我们能够在他对于风险的行为中得到一种有偏性。在附录Ⅱ中我们将证明，如果研发项目类限制在我们前面已经讨论过的种类中，那么垄断者将在它的研发中承担过低的风险程度，而且在研发中投资过

小。但是，此时即使可获得的研发项目类型具有稍低的期望成本降低幅度（$c-c^*$），和较高的风险（即α变大），这一点仍然是对的。在这种情形下，如果该行业是由一个私人垄断者所控制，那么我们将观察到，平均来说该行业的成本降低幅度比被社会控制时的成本降低幅度更大。这个结果的关键在于，虽然成本降低是一件“好事情”，平均来说私人垄断者反而会使成本降低得更多，虽然其研发支出小于社会最优水平。

这些纯粹的经济环境都是最容易分析的情形。在自由进入条件下，分析显得不同寻常地复杂。但是在寡头竞争环境中，一个生产者的市场控制力会随着它相对于对手的成本优势的增加而增加。从而我们可能有如下结论，竞争可能鼓励生产者在过程创新中承担过度的风险。在存在对该行业自由进入的条件下，我们不能明确地回答这个问题。

四、市场结构与研发速度

我们现在关注研发的速度，而且我们将继续在过程创新的环境中讨论问题。我们回到第一部分的基本模型，并且假设c为当前实际中存在的最好的生产技术的单位生产成本。就像我们在第一、二部分中的做法一样，我们做出简化，假设研发是高度的具有目的导向性的。现在研发的内容为解决一系列能够使得某种商品的单位生产成本降为c^*（$c^*<c$）的问题。而且这里不存在不确定性。[17]如果一个研发单位在$t=0$时刻投资x，那么它将在时刻$T(x)$解决所有的问题，其中$T'(x)<0$且$T''(x)>0$，当$x\to\infty$时$T(x)\to 0$，而当$x\to 0$时$T(x)\to\infty$（见图5）。[18]

首先考虑竞争性条件。我们假设当前该商品的市场价格为c。所有人都知道，最早发明技术的那家公司将被奖励在一定年限内享有专利权。如果有多个胜者，那么它们分享该专利。我们并不需要像在第一、二部分那样，明确假设在后一种情形中那些胜者是否通过组成卡特尔以达到勾结的目的，或者是否以古诺的方式进行博弈。所有我们需要做出的规定就是，在专利有效期间，每一个胜者都能享有一个正的利润流。从而企业都是很有动机从事研发的。我们假设每一家企业都知道其他企业的研发战略。

---

17 关于在研发成功时间方面存在不确定性的模型曾在Stiglitz（1970）；Dasgupta，Heal and Majumdar（1977）；Kamien and Schwartz（1978）；Dasgupta and Stiglitz（1978）等论文中有过讨论。

18 即所有的潜在研究单位的研发能力都是一样的。做这个假设仅仅是为了说明简单，放松这个假设的方式是很明显的。

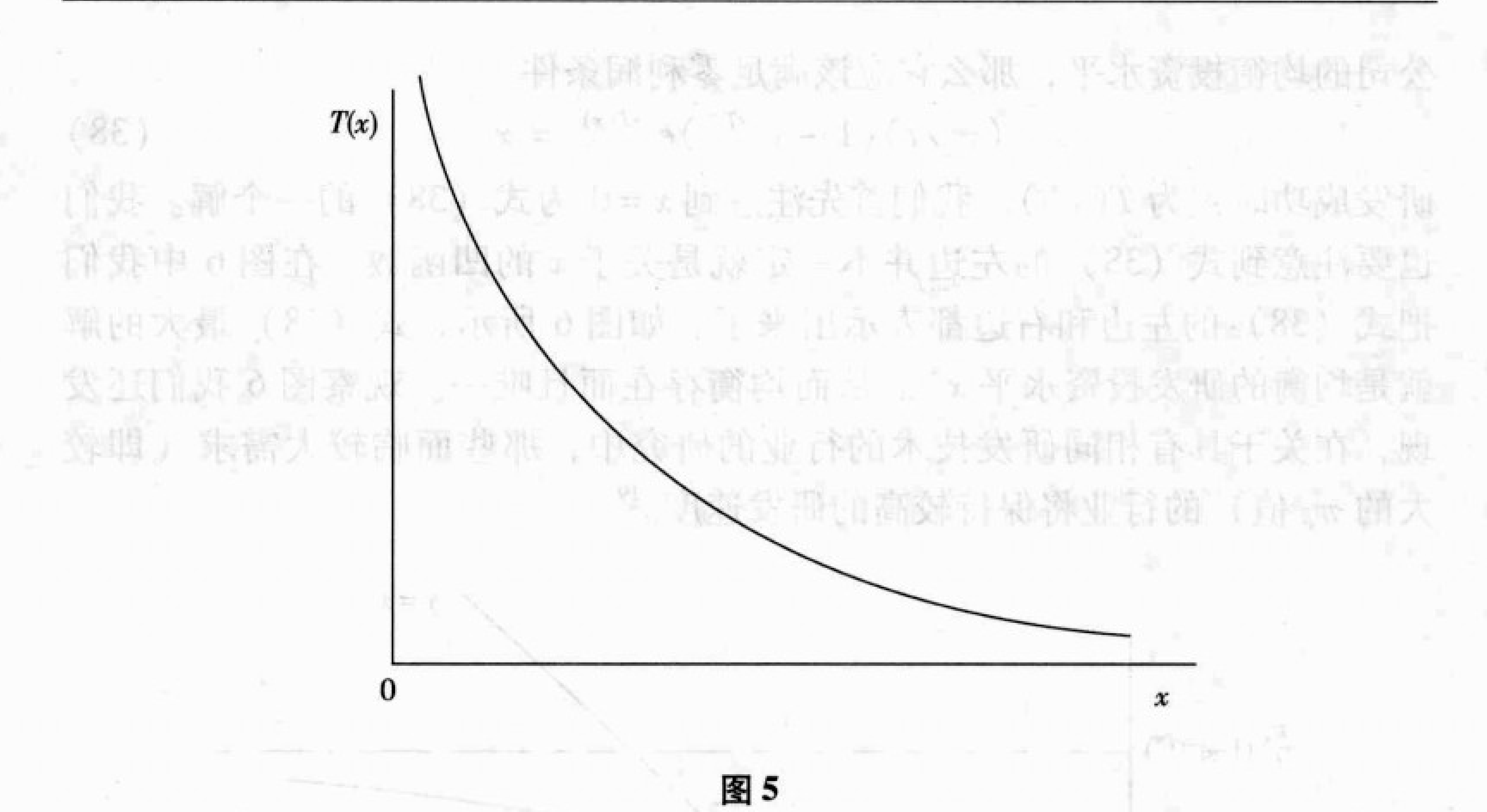


图 5

企业被假设为能够以一个竞争性利率 $r(r>0)$ 自由借贷来为其研发融资。既然根据假设 $T'(x)<0$，那么每一家从事研发的企业都只会建立一个研究机构。跟前面的做法一样，为了得到博弈的均衡，我们假设，关于其他从事研发企业的行为，每一家企业都接受古诺假设。但是，我们希望能够反映研发的企业害怕潜在进入者的进入威胁这一事实。所以，我们假设，关于潜在进入者的行为，每一家公司都接受斯塔尔伯格（Stackelberg）的假设，换句话说，它们将考虑潜在进入者的反应函数。后一条假设意味着，在对于研发活动的自由进入条件下，均衡时每一家企业将获得零现值的利润流，而前一条假设将意味着只有将来的胜者才从事研发。

现在，假设在一个潜在的均衡有不止一家公司处于经营状态。因而它们都投资同样的数目。但是这不可能成为一个均衡。因为在关于其他参与者的古诺假设下，其中任何一家企业都可以通过边际提高投资额，使自己成为唯一的赢家，从而使自己利润流的现值有一个非连续的跳跃。所以我们得到在自由进入条件下，在均衡时最多只能有一家公司从事研发，而且其利润流的净现值为零。

这个结果告诉我们，只有一家公司从事研发活动这个事实本身并不能成为该领域竞争很少的证据。对于我们现在这个模型来说，竞争是很激烈的。这家唯一的从事研发的公司通过提高研发支出，将它本身的研发速度提高至一个新的高度来阻止进入。

现在，我们仍然需要计算均衡的支出水平。令 $T^*(>0)$ 为专利的有效期长度。我们现在回顾一下第一部分引入的概念。如果 x^* 为那一家唯一的

公司的均衡投资水平，那么它应该满足零利润条件

$$(\pi_c/r)(1-e^{-rT^*})e^{-rT(x)} = x \tag{38}$$

研发成功时刻为 $T(x^*)$。我们首先注意到 $x=0$ 为式（38）的一个解。我们也要注意到式（38）的左边并不一定就是关于 x 的凹函数。在图 6 中我们把式（38）的左边和右边都表示出来了。如图 6 所示，式（38）最大的解就是均衡的研发投资水平 x^*。从而均衡存在而且唯一。观察图 6 我们还发现，在关于具有相同研发技术的行业的研究中，那些面临较大需求（即较大的 π_c 值）的行业将保持较高的研发速度。[19]

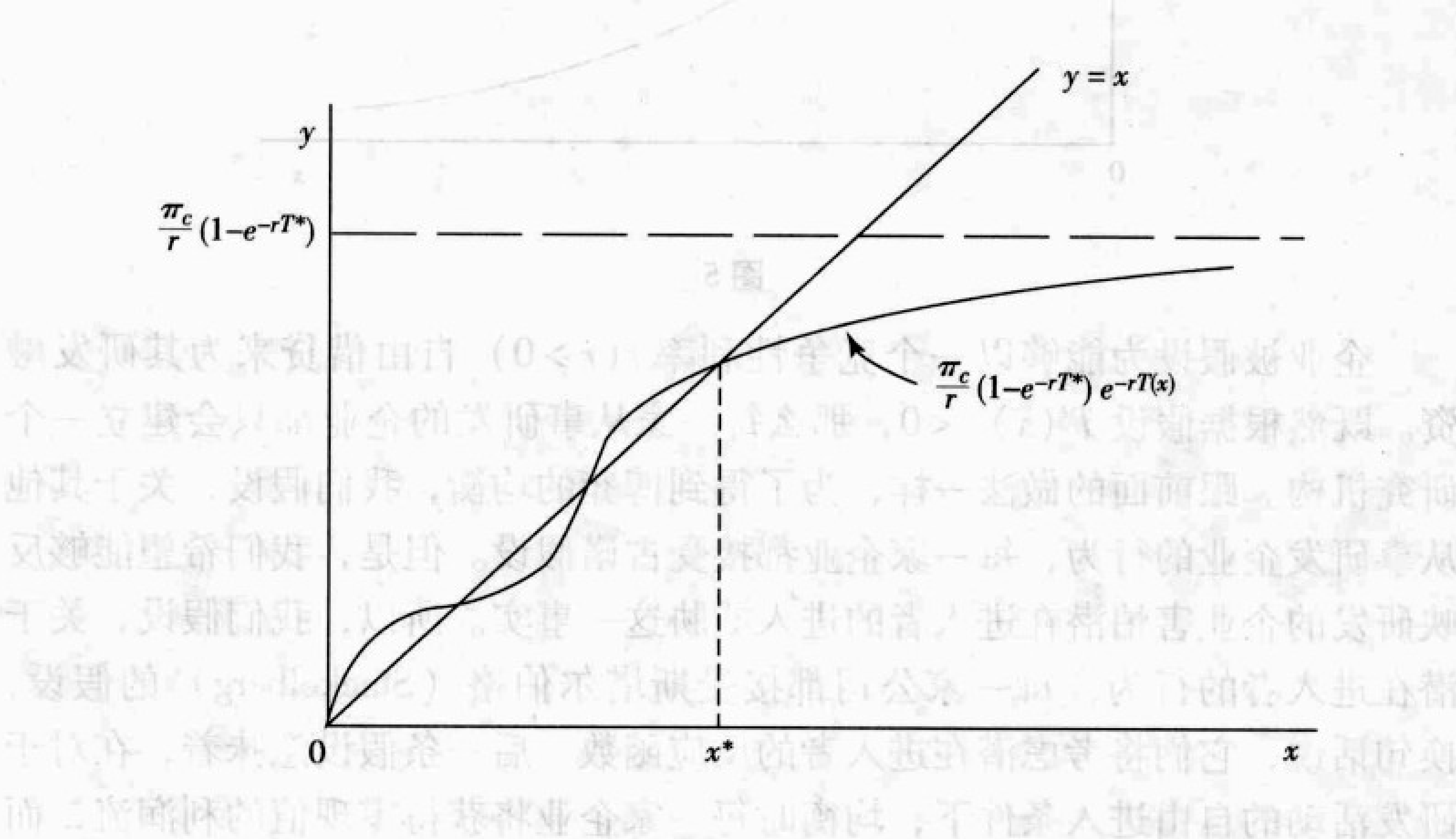


图 6

我们现在比较一下竞争性结果与社会最优的研发速度。我们假设 r 为合适的折旧率。从而社会计划者关注如何选择 x 以最大化

$$\left[\frac{\pi_s}{r}e^{-rT(x)} - x\right] \tag{39}$$

19　为了获得关于大小关系的一个直观感觉，我们注意到在大公司中一个研发项目一般预期在 4～5 年内完成，而且许多这样的项目都包含适当的风险，而且在许多情况下估计的成功概率都要超过 0.8（Mansfield，1967）。关于风险研发项目的明显例子出现在制药行业中（Schwartzman，1977）。关于市场结构、研发中的风险分担、研发速度等之间的关系的理论分析请见 Dasgupta and Stiglitz（1977，1980）。

再一次地，我们注意到式（39）一般来说不是关于 x 的凹函数。但是，给定我们关于 $T(x)$ 的假设，马上就有式（39）存在最大值。令 x_s 为最优的研发支出。从而，给定其为正，那么它满足成本—收益规则

$$-T'(x)\pi_s e^{-rT(x)}=1 \tag{40}$$

回顾一下 $\pi_s>\pi_c$（不等式（2））。但是根据式（38）和式（40），我们可以很容易地得到 $x^*>x_s$。从而，由于竞争的威胁使得唯一一家从事研发活动的公司被迫在研发方面过度投入。市场竞争鼓励过快的研发速度。

如果我们假设所有的厂商在解决所有为了得到 c^* 生产技术的必要问题方面必须采取相同的步骤（换句话说，即如果所有的研发单位都面临相同的决策树）的特殊情形下，我们引入不确定性，本部分的中心结论——在自由进入的条件下，我们这个模型的一个纳什—古诺均衡最多能容纳一个企业——仍然能够保持。这意味着给定其他企业的研发活动的速度，某家企业总可以通过选择一个足够快的研发速度，以保证它是第一家发明该项技术的企业，虽然它仍然不确切知道在哪一天研究能够最终被它的研发机构所完成。在这种形式的不确定性条件下，我们马上可以得到在均衡时，最多只有一家企业会对研发进行投入。

现在，该假设将使得各企业研发战略间的相互依赖性不再适当。在另一篇文章（Dasgupta and Stiglitz，1978）中我们走到了另一个极端，假设它们之间完全独立。在这样一种情形下，显然均衡可以包含多家企业在研发方面竞争；因为没有哪一家企业能够保证自己就是胜者。

因此，关键在这里。均衡时在研发方面竞争的企业的数目决定的一个关键因素就是，各企业之间成功概率的相关程度。一般来说，人们会认为相关程度较高时企业数目会少些。但是从事研发的企业数目并不是一个衡量竞争程度的合适的指标。正如我们前面所看到的，竞争有可能导致只有一家公司从事研发方面的投资，而且投资过度，并且导致过快的研发速度。

五、结论

现在，有相当多的经验文献在研究研发活动的方方面面，而且它们企图把研发活动与其所处的行业结构相联系起来。因此我们很惊讶地发现关于这个领域的理论研究寥寥无几。到现在为止，这些为数不多的理论研究的大部分，还都着重关注在一个给定的环境中单个企业在研发方面的决策分析。为了向一些经验研究成果靠拢，也为了获得关于这个问题的一个较完全的理解，很有必要将我们的工作拓展到对单个公司的分析之外，考虑一个相互作用的

公司的集合。如果有人对长期也很感兴趣，那么就还有必要在分析中假设一个内生的市场结构。这就形成了一个博弈的环境，在本文，我们就尝试着在一些简单假设构造的环境中研究产业组织理论中的一些很传统的问题。在研究市场经济中的研发活动的整个过程中，我们都是利用博弈均衡的概念来分析的。当然，众所周知，这种类型的均衡的存在及其性质取决于我们假设博弈参与者采取什么样的行动，以及博弈参与者对其他博弈参与者对于自己的决策反应的猜想。在本文中，我们假设企业选择数量（产出水平和研发支出水平）。关于猜想到现在为止最容易分析的就是古诺类型的猜想，即其他参与者并不对该参与者的决策做出反应。在本文的绝大部分中，我们都是假设这种类型的猜想。这样的假设虽然在一定条件下（例如，第二部分中的模型中）是合理的，但是在很多其他条件下却并不是很合理。例如，读者可以检查一下，在第四部分中的模型中，如果所有的企业都持有关于其他企业的古诺猜想，那么该模型就不存在均衡。所以我们在第四部分中的分析中就采用了一种类似混合的假设；“a－priori”的假设，这是有其合理性的。然而，在关于公司的更复杂的猜想条件下，考察寡头竞争均衡的特征，以及考察这些特征如何随着关于公司所持猜想的变化而变化显然是必要的。[20]

在产业组织理论文献中，被讨论频率最高的问题就是公司规模和研发速度之间的关系。在本文的导言和第二部分中，我们讨论了该问题的一些方面。现在，如果公司被强迫使用自有资金为其研发活动融资的话，那么显然我们有产业集中度和研发活动正相关的结论。第二部分中的模型并没有引入时间，其结果与内部融资假说是一致的。但是，在第四部分，为了说明要得到一般均衡时，只有有限家企业从事研发活动的结论，我们并不需要内部融资假说，我们假设了一个发育完全的资本市场。事实上，正如我们所说明的，为了与某些经验研究结果相一致，我们也不需要假设从公司的角度而言的非经典的目标。通篇我们都假设在企业从事博弈的市场经济环境中具有完全信息。在这样的环境下，除了假设企业以利润最大化为目的，其他的假设很难说得通。但是如果假设关于企业的不完全信息，那么情况就会有些不同；而且最近关于这方面的一个很有意思的探索（Winter，1971；Nelson，Winter，Schuette，1976）就是除此之外，还假设厂商在解复杂的最大化问题方面的有限能力。在“有限理性”的条件下，厂商被假设为遵循一定的

20　关于存在进入障碍条件下，且侧重于假设公司持有其他类型猜想的寡头竞争均衡的一个重要的分析请见 Marschak and Selton（1974）。

"满意"行动路径，如果现行的行动路径不再能够产生满意的结果，那么就在局部寻找改进的机会。

除了有些看来可能还比较有意思的细节外，我们分析的主要结论主要有以下一些：

（1）即使当有人认为市场结构是内生的时，理论分析与当行业的市场集中度比较小时，行业范围内的研发活动与行业集中度是正相关的经验结果也是一致的（第二部分）。

（2）高的市场集中度本身并不能作为行业缺乏有效竞争的证据（第二部分）。

（3）当自由进入条件下行业集中度较小时，每一家公司的研发努力（从而生产成本降低）与集中度一般来说是正相关的，而且此时集中度是与垄断程度正相关的（第二部分）。

（4）在一个市场经济中的研发支出和每一家公司的研发支出都是与市场规模正相关的。如果需求是富有弹性的，那么它们随着研发技术成本的提高而下降，如果需求是缺乏弹性的，那么结论相反（第二部分）。

（5）当存在进入障碍时，一个市场经济中公司数目的增加将导致平均每家公司的研发支出减少，虽然该行业的总产出水平会提高，从而垄断程度将会降低（第二部分）。

（6）即使在自由进入的条件下，我们仍然有在市场经济中行业的生产成本降低幅度要低于社会最优水平，但是

（7）如果需求是高度的缺乏弹性，自由进入条件下一个行业总的研发支出要超过社会最优水平（第二部分）。上述第（6）点和第（7）点意味着：

（8）虽然市场经济中行业的生产成本降低幅度较低，但是市场经济条件下整个行业的总研发支出要超过社会最优水平，因此从这个意义上说，市场经济将导致企业研发努力的过度重复。特别是，一个行业可能同时均有很低的行业集中度（存在很多家企业）和很高程度的浪费。

（9）在过程创新情形下，一个好的结果就是鼓励对风险研发项目的投资，即使整个社会是风险规避的（第三部分）。

（10）一个纯粹的垄断者（受到进入门槛的保护）不具备充分的动机去：(a）进行研发投资（第二部分）；（b）从事风险研发项目（第三部分和附录Ⅱ）。

（11）既然当一家公司相对于其对手的成本优势增加时，其市场控制力也就增加了，那么我们就有，市场鼓励企业从事风险过高的研发项目（第

三部分）。

（12）如果某发明的绝大部分奖赏都归于首先做出该发明的企业，那么在一定程度上说，企业所承担的风险是正相关的，而且竞争的压力将保证，均衡时只有为数不多的几家企业从事研发活动，而在极端的情况下，只有一家企业能够从事研发活动（第四部分）。

（13）但是，我们所观察到的只有少数几家企业从事研发活动的事实，本身并不能作为市场经济导致过少的研发的证据。特别是：

（14）竞争的压力将导致过快的研发速度（第四部分），而且一般来说：

（15）并不存在市场经济倾向于产生过少的知识的结论。

附录 I

我们将在这里给出对于第二部分中的模型的一个自由进入的对称均衡存在的条件，其中的 $p(Q)$ 和 $c(x)$ 满足文中的式（7）和式（9）。为了证明这个结论，我们将正文所遵循的顺序倒过来，首先假设存在进入障碍，从而企业数目 n 是外生给定的。

考虑两个一阶条件式（30）和式（31）的解式（32）和式（33）。为了让它们有实值解，我们必须假设 $n>\epsilon$。让我们这样假设。注意到当假设每一家其他的企业都选择 Q^*/n 作为产出水平时，（x^*，Q^*/n）为关于该有代表性的厂商利润最大化一阶条件的唯一内点解。我们也必须给出当该有代表性企业选择（x^*，Q^*/n）时，其赚取非负利润的条件。

令 $\pi(x^*, Q^*/n)$ 表示利润水平。那么 $\pi(x^*, Q^*/n) = [p(Q^*) - c(x^*)]Q^*/n - x^*$。要使这个值非负，那么其充分必要条件为

$$[p(Q^*) - c(x^*)]Q^*/nx^* \geqslant 1 \tag{A1}$$

如果我们现在利用函数形式（7），式(9) 连同式（32）和式（33），那么常规的计算就可以得到

$$[p(Q^*) - c(x^*)]Q^*/nx^* = \epsilon/\alpha(n-\epsilon) \tag{A2}$$

注意到当 $n>\epsilon$ 时，式（A2）的右边为关于 n 的递减函数，而且当 $n=\epsilon(1+\alpha)/\alpha$ 时，其值为 1。结合式（A1）和式（A2）我们得到

$$\pi(x^*, Q^*/n) \geqslant 0 \text{ 如果 } \epsilon < n \leqslant \epsilon(1+\alpha)/\alpha \tag{A3}$$

从而，式（A3）要求我们假设 $\epsilon<n\leqslant\epsilon(1+\alpha)/\alpha$。

我们还需要找到对于企业 i，其利润函数 $\{p[(n-1)Q^*/n+Q_i] - c(x_i)\}Q_i - x_i$ 在点 $x_i=x^*$，$Q_i=Q^*/n$ 点附近为凹函数的条件。事实上，如果函数 $p(Q)$ 和 $c(x)$ 具有式（7）和式（9）所示的函数形式，而且 x^* 和

Q^*/n 由式（32）和式（33）给出，那么我们可以验证，如果 $\epsilon > \alpha(1-\epsilon)$（即 $\epsilon(1+\alpha)/\alpha > 1$），那么该有代表性的公司在 x^* 和 Q^*/n 处满足二阶条件。所以，如果每一家其他的企业都选择（x^*，Q^*/n），那么（x^*，Q^*/n）就是该具有代表性的企业的利润最大化选择（图 7）。从而我们证明了

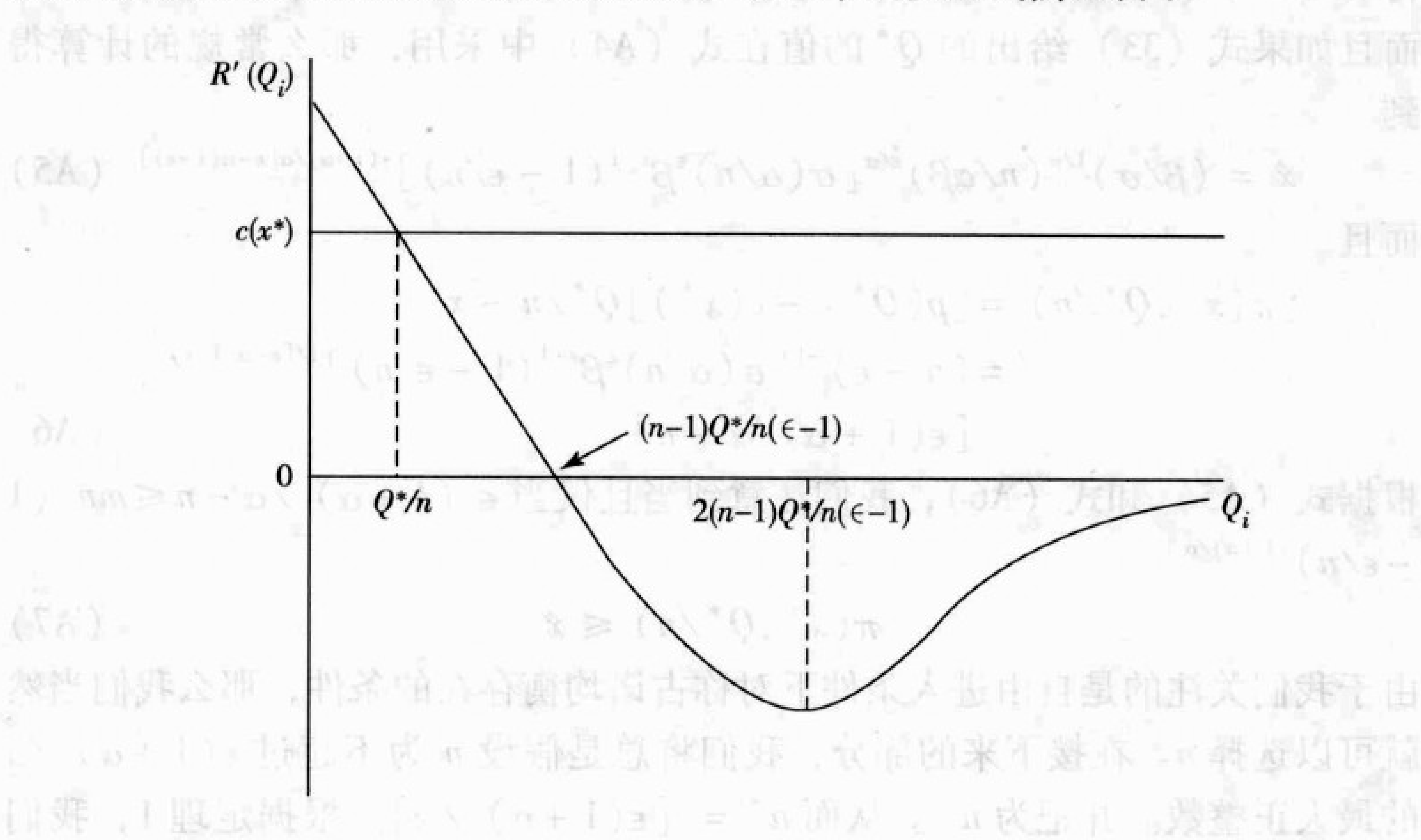


图 7　当收入为 $R(Q_i) = \sigma[Q_i + (n-1)Q^*/n]^{-\epsilon}Q_i$ 时企业 i 的边际收入曲线 $R'(Q_i)$ 其中 Q^* 为均衡的行为产出，$\epsilon>1$，且 $n>\epsilon$。x^* 为四个企业的均衡研发支出

定理 1：如果 $p(Q)$ 和 $c(x)$ 分别满足式（7）和式（9），而且 n 是一个正整数，那么这 n 家公司中存在一个对称的古诺均衡，如果

$$(\text{i})\epsilon < n \leqslant \epsilon(1+\alpha)/\alpha \text{ 和 } (\text{ii})\epsilon(1+\alpha)/\alpha > 1$$

满足。

现在，显然当 $\epsilon > 1$ 时 $\epsilon(1+\alpha)/\alpha > 1$。从而我们所证明的就是，如果 $\epsilon < n \leqslant \epsilon(1+\alpha)/\alpha$，那么 n 家企业的古诺对称均衡将存在，即使市场需求是全局缺乏弹性的。

我们下面转入在自由进入条件下的对称古诺均衡存在性条件的讨论。假设行业中存在 n 家企业，而且假设从此不存在对进入的障碍。然后，令（x^*，Q^*/n）为在一个对称的古诺均衡中一个有代表性的企业的行动路径选择。那么此时产品的市场价格为 $p(Q^*)$。现在定义下式的解为 $\hat{x}$：

$$p(Q^*) = c(x) \tag{A4}$$

$\hat{x}$ 就类似一个进入成本。此时显然存在 $m > 0$，独立于 n，使得，如果

$\pi(x^*,\ Q^*/n)\ \leqslant m\hat{x}$，则其他的企业将会发现，即使不存在进入障碍，进入该市场也不再有利可图。所以，事实上，如果 n 有一个这样的值使得 $\pi(x^*,\ Q^*/n)\ \leqslant m\hat{x}$，那么正文中的式（19）将被满足。我们下面给出使得 $\pi\ (x^*,\ Q^*/n)\ \leqslant m\hat{x}$ 的条件。$p(Q)$ 和 $c(x)$ 满足式（7）和式（9），而且如果式（33）给出的 Q^* 的值在式（A4）中采用，那么常规的计算得到

$$\hat{x} = (\beta/\sigma)^{1/\alpha}(n/\alpha\beta)^{\epsilon/\alpha}[\sigma(\alpha/n)^{\epsilon}\beta^{\epsilon-1}(1-\epsilon/n)]^{\epsilon(1+\alpha)/\alpha[\epsilon-\alpha(1-\epsilon)]} \quad (A5)$$

而且

$$\begin{aligned}\pi(x^*,Q^*/n) &= [p(Q^*)-c(x^*)]Q^*/n - x^* \\ &= (n-\epsilon)^{-1}[\sigma(\alpha/n)^{\epsilon}\beta^{\epsilon-1}(1-\epsilon/n)]^{1/[\epsilon-\alpha(1-\epsilon)]} \\ &\quad [\epsilon(1+\alpha)/\alpha - n]\end{aligned} \quad (A6)$$

根据式（A5）和式（A6），我们注意到当且仅当 $\epsilon\ (1+\alpha)\ /\alpha - n \leqslant mn\ (1-\epsilon/n)^{(1+\alpha)/\alpha}$

$$\pi(x^*,Q^*/n) \leqslant \hat{x} \quad (A7)$$

由于我们关注的是自由进入条件下对称古诺均衡存在的条件，那么我们当然就可以选择 n。在接下来的部分，我们将总是假设 n 为不超过 $\epsilon(1+\alpha)\ /\alpha$ 的最大正整数。并记为 n^*。从而 $n^* = [\epsilon(1+\alpha)\ /\alpha]$。根据定理 1，我们可以注意到

定理 2：[21] *如果 $p(Q)$ 和 $c(x)$ 分别满足式（7）和式（9），那么一个自由进入的对称古诺均衡存在，如果*

$$(\mathrm{i})\epsilon(1+\alpha)/\alpha > 1, (\mathrm{ii})n^* = [\epsilon(1+\alpha)/\alpha] > \epsilon$$

且

$$(\mathrm{iii})\epsilon(1+\alpha)/\alpha - n^* \leqslant mn^*(1-\epsilon/n^*)^{(1+\alpha)/\alpha}$$

该定理使得保证自由进入的对称古诺均衡存在的经济参数间的关系显式化了。特别是，我们也应当注意到，如果对于任意给定的 ϵ 值，α 被选择得充分地小，而且同样，对于任意给定的 α 值，ϵ 值被选择得充分地大，那么定理 2 的条件就能够得到满足。那就是说，如果 ϵ/α 足够大，那么均衡将会存在。这就验证了正文中的结论。

我们还需要找到，在什么条件下零利润条件（21）的使用是一个良好的近似。事实上我们已经给出了这个条件。我们所要找的正是使得

---

21 Novshek（1977）独立地在一个与本文不尽相同的环境中证明了一个与此相似的结论。我们要感谢 Morton Kamien 提醒我们注意到了 Novshek 的工作。

$\pi(x^*, Q^*/n^*)/x^*$ "较小"的条件。从而条件（A2）告诉我们，对于文中式（7）和式（9）给出的具体化形式，这就相当于寻找使得 $\epsilon/\alpha(n^*-\epsilon)-1$ 可忽略的条件。既然根据构造 $n^* = [\epsilon(1+\alpha)/\alpha]$，那么我们就可以给出使零利润条件成为一个好的近似的（α, ϵ）的取值范围。特别地，我们注意到，如果 ϵ/α "较大"，那么它就是一个好的近似。在这种情况下当然 n^* 就"较大"。

附录Ⅱ

在这一部分，我们比较纯粹的垄断情况下的风险承担度和社会最优的风险承担度。为了这个目标，我们假设研究项目族由式（34）表达，而且假设对于 $\alpha \geqslant 0$，$R(\alpha) > 0$，$R'(\alpha) > 0$ 且 $R''(\alpha) > 0$。设市场需求函数由社会效用函数式（6）生成。从而需求函数为

$$p(Q) = \sigma Q^{-\epsilon}(1 > \epsilon > 0) \tag{A8}$$

令 $V_m(c)$ 表示当 c 为单位生产成本时垄断者的最大销售利润。那么很容易地，我们可以验证

$$V_m(c) = \epsilon\sigma^{1/\epsilon}(1-\epsilon)^{(1-\epsilon)/\epsilon}c^{-(1-\epsilon)/\epsilon} \tag{A9}$$

如果垄断者选择项目 $-\alpha$，那么它的下一期期望利润为

$$h(\alpha)V_m[c(\alpha)] + [1-h(\alpha)]V_m(c) - R(\alpha) \tag{A10}$$

其中 $V_m(c)$ 由式（A9）给出，记

$$A \equiv \epsilon^{1/\epsilon}(1-\epsilon)^{(1-\epsilon)/\epsilon}, g(\alpha) \equiv \{h(\alpha)c(\alpha)^{-(1-\epsilon)/\epsilon} + [1-h(\alpha)]c^{-(1-\epsilon)/\epsilon}\}$$

那么式（A10）可以表示成

$$Ag(\alpha) - R(\alpha) \tag{A11}$$

为了分析的简单，我们现在假设研究项目族使得函数 $g(\alpha)$ 为关于 α 的凹函数。那么给定（期望）利润最大化的研究项目严格落在 $\alpha=0$ 和 $\alpha=\bar{\alpha}$ 之间，那么垄断者将选择的 α 为下式的解

$$Ag'(\alpha) = R'(\alpha) \tag{A12}$$

令 α_m 为式（A12）的（唯一）解。

现在，我们转而分析社会中央计划者。他关注于社会收益式（36）。既然社会效用函数具有式（6）所示的形式，那么 $V(c)$ 满足式（8）。记 $B \equiv \varepsilon\sigma^{1/\epsilon}(1-\epsilon)^{-1}$。那么式（36）可以写成

$$Bg(\alpha) - R(\alpha) \tag{A13}$$

因为根据假设式（A13）关于 α 严格凹，那么计划者将也会愿意从事一些基本的研究项目。令 α_s（$0<\alpha_s<\bar{\alpha}$）为社会最优的选择项目。那么 α_s 必

须满足下式

$$Bg'(\alpha) = R'(\alpha) \tag{A14}$$

但是 $B > A$。比较式（A12）和式（A14），我们得到 $\alpha_s > \alpha_m$。这就证明了本文中的结论。

参考文献

Arrow, K. J. (1962). 'Economic welfare and the allocation of resources for invention.' In *The Rate and Direction of Inventive Activity: Economic and Social Factors* (ed. R. Nelson). (NBER) Princeton University Press.

Barzel, Y. (1968). 'Optimal timing of innovation.' *Review of Economics and Statistics* (August), pp. 348–55.

Dasgupta, P., Heal, G. M. and Majumdar, M. (1977). 'Resource depletion and research and development.' In *Frontiers of Quantitative Economics*, vol. IIIB (ed M. Intrilligator). Amsterdam: North Holland.

—— and Stiglitz, J. E. (1977). 'Market structure and research and development.' London School of Economics (mimeo).

—— —— (1980). 'Uncertainty, market structure, and the speed of research.' *Bell Journal of Economics* Spring.

Evenson, R. and Kieslev, Y. (1975). 'A model of technological research.' *Journal of Political Economy.*

Kalecki, M. (1954). *Theory of Economic Dynamics*. London: George Allen and Unwin.

Kamien, M. and Schwartz, N. (1972). 'Timing of innovations under rivalry.' *Econometrica*, 40, no. 1, pp. 43–60.

—— —— (1975). 'Market structure and innovation: a survey.' *Journal of Economic Literature*, vol. 13, no. 1, pp. 1–37.

—— —— (1978). 'Optimal exhaustible resource depletion with endogenous technical change.' *Review of Economic Studies*, vol. 45, pp. 179–96.

Levin, R. C. (1978). 'Technical change, barriers to entry and market structure.' *Economica*, vol. 45, no. 180, pp. 347–62.

Loury, G. (1977). 'Market structure and innovation.' NorthWestern University, Center for Mathematical Studies in Economics and Management Science, Discussion Paper, no. 256.

Mansfield, E. (1967). *Econometric Studies of Industrial Research and Technological Innovation*. New York: W. W. Norton & Co.

Marschak, T. and Selton, R. (1974). *General Equilibrium with Price-Making Firms*. Berlin: Springer-Verlag.

Nelson, R. R., Peck, M. J. and Kalachek, E. D. (1967). *Technology, Economic Growth and Public Policy.* Washington, D.C.: Brookings Institution.

—— and Winter, S. (1977), 'Forces generating and limiting concentration under Schumpeterian competition.' *Bell Journal of Economics*, vol. 9, no. 2 (Autumn), pp. 524–48.

—— —— and Schuette, H. L. (1976). 'Technical change in an evolutionary model.' *Quarterly Journal of Economics*, vol. 90, no. 1, pp. 90–118.

Novshek, W. (1977). 'Nash–Cournot equilibrium with entry.' Mimeo. Princeton University.

Scherer, F. M. (1970). *Industrial Market Structure and Economic Performance*. Chicago: Rand McNally.

Schmookler, J. (1962). 'Economic sources of inventive activity.' *Journal of Economic History*, vol. 22 (March).

Schumpeter, J. (1947). *Capitalism, Socialism and Democracy* (2nd edn.). London: Allen and Unwin.

Schwartzman, D. (1977). *Innovation in the Pharmaceutical Industry*. Baltimore: John Hopkins University Press.

Stiglitz, J. E. (1970). 'Perfect and imperfect capital markets.' Yale University, mimeo. Paper presented to the Winter meeting of the Econometric Society (1970).

Winter, S. (1971). 'Satisficing, selection, and the innovating remnant.' *Quarterly Journal of Economics.*

不确定性、产业结构与研发速度*

本文通过比较分析研究的速度、独立研究实验室的数量和企业承担的风险水平，探讨了研发活动中竞争的性质和后果，以及研发活动中的竞争与产品市场上的竞争的关系，并得到了以下结论：（相对于垄断而言）目前产品市场上的竞争会降低创新水平；研发活动中的竞争会提高创新水平，甚至可能会使创新水平超过社会最优水平。在一定条件下，创新会使垄断企业领先于潜在的竞争者，从而使垄断得以维持。在市场均衡中，研究速度可能会超过最优水平，而企业承担的风险水平又可能是不充分的。

一、引言

在现代资本主义经济中，竞争有很多种方式。不但有价格竞争——这是传统的微观经济理论关注的重点——而且还有产品竞争（Spence，1976；Dixit and Stiglitz，1977；Salop，1977）。在本文，我们将关注第三个领域。在此领域中，竞争发生在发明与创新活动中。虽然在产品竞争理论中，可能的产品集合是固定的，而且每一种可能的产品的生产技术是给定的，但是发明与创新的理论关注的焦点是，新产品和新生产技术的发展。尽管熊彼特在1947年就对这一类竞争进行了富有启发性的描述，但是，迄今为止，有关分析此类竞争的理论探索依然十分有限。

---

* "Uncertainty, Industrial Structure, and the Speed of R&D", with Partha Dasgupta, *The Bell Journal of Economics*, Vol. 11, No. 1 (Spring, 1980), pp. 1 – 28. 两位作者当时分属伦敦经济学院和普林斯顿大学。

本文是以一篇特约论文（Dasgupta and Stiglitz，1977）的第九部分和第十部分为基础修改而成的。在1977年8月27日~9月3日于东京举行的国际经济协会关于“经济增长与资源”学术大会（the World Congress of the *International Economic Association on Economic Growth and Resources*）上，我们受邀宣读了该论文。国家科学基金会资助了该项研究。在1978年秋准备出版目前这一版本时，达斯古普塔为贾瓦哈拉尔·尼赫鲁大学国际研究学院和德里经济学院（德里）的访问教授，斯蒂格利茨是Mathematica的奥斯卡·摩根斯坦杰出学者和普林斯顿高级研究学院的访问教授。保罗·戴维（Paul David）、桑福德·格罗斯曼（Sanford Grossman）、阿肖克·古哈（Ashok Guha）、埃德温·曼斯菲尔德（Edwin Mansfield）、大森隆史（Takashi Omari）、罗勃·波特（Rob Porter）、编辑委员会，尤其是理查德·吉尔伯特（Richard Gilbert）等人深刻的评论让我们获益良多。本文中的一些思想，就来自于与他们进行的广泛而深入的讨论。

传统的竞争均衡模型并不考虑各个产业都会经历的“持续不断的创造性破坏浪潮”的情形。如果存在研发支出，那么在考察诸如市场结构与技术进步速度之间的关系这类基本问题时，传统模型所采用的假设（例如，技术的凸性）将肯定得不到满足。[1]

自从熊彼特的著作问世，对某一产业的结构与该产业中的研发活动的关系的探讨，在很大程度上，都局限于研究产业集中度对研发支出的影响这方面。在由谢勒（Scherer，1970）以及卡曼和施瓦茨（Kamien and Schwartz，1975）全面综述的经验文献中，绝大部分都在探讨产业集中度与研发努力之间的因果关系。当然，对于短期分析而言，非常明智的做法是，先将产业集中度视为一个数据，进而去寻求因果关系。但是，在长期，一个给定产业的结构本身必定是内生的。虽然我们期望研发支出与产业结构之间存在一定的关系，但是从理论上讲，研发支出和产业结构都取决于一些更为基本的因素，例如，研究的技术、需求条件以及资本市场的特性。[2] 尼尔森和温特（Nelson and Winter，1975）最近的一项模拟研究就体现了这种思想。

要想将产业结构和研发支出与这些基本因素联系起来，就不但要去分析产品市场的竞争方式，而且还要去分析研发活动中的竞争方式。同时，这也意味着，虽然研发活动中的竞争不是完全竞争——这是熊彼特十分强调的一点——但是，它同样也不是通常的垄断竞争。在垄断竞争理论中，十分接近的（但也是不完全的）替代品的存在，会使一个给定企业的市场势力受到限制。与之相对比，在我们所考虑的环境中，在一个给定的时点，一家生产普通商品的企业也许会面对无效的竞争。其原因可能是，该企业享有一项生产该种普通商品的专利权，而为了获得这项专利权，该企业在过去花费了许多成本。[3] 不过，当最初的企业继续开展研发活动以免其垄断地位受到侵蚀时，其他企业也会通过承担研发支出来为分享未来的市场份额而展开竞争。

1　参见斯蒂格利茨对知识特征进行的广泛而深入的讨论。知识所具有的特征会使那些对普通商品的生产所做出的常规假设，无法适用于研发活动。同时，参见 Dasgupta and Heal（1971）。

2　当然，就像加尔布雷斯（Galbraith，1973）最近所强调的那样，一个更为完整的理论甚至应该将需求条件内生化。

3．然而，很重要的一点是要注意到，即使在短期，而且在一家企业垄断现有技术的条件下，若存在研发竞争，则传统的利润最大化垄断企业模型也许就不适用了。原因在于，该企业今天（与销售、生产能力等有关）的行动，可能会产生重要的进入阻碍作用。企业会考虑到这一点的。例如，参见 Dasgupta，Gilbert and Stiglitz（1979），Gilbert and Stiglitz（1979），Spence（1977），Salop（1979）。

在这一环境中，竞争的关键方式就是研发。[4]

在本文及更早的一篇文章（Dasgupta and Stiglitz，1980）中，我们主要分析了产业结构与研发活动不同特征之间的关系。在达斯古普塔和斯蒂格利茨（Dasgupta and Stiglitz，1980）的分析中，我们注意到，在过程创新模型中，只要市场均衡状态下的产业集中度不是非常高，那么在产业的研发支出与集中度之间就存在正相关关系，而在经验文献中通常会将这种相关关系视为因果关系（Scherer，1970）。同时，我们也注意到，我们可以推知在市场经济中，研发活动会出现过度重复的现象。虽然在某种意义上，每个企业所承担的研发活动要低于社会最优水平，但是在市场均衡状态下，开展研发活动的企业数目会维持在一个不恰当的水平上。因此，产业范围内的研发支出就过多了。在本文中，我们关注另外两个问题：研发活动发生的速度，以及市场经济中开展研发活动的企业所能承担的风险程度。

我们从两个极端的情形中展开本文的分析。一是（第三部分），我们假设，不同研究单位所面临的不确定性是完全相关的，下面我们会准确地说明其含义。二是（第四部分和第五部分），我们假设，各研究单位所面临的不确定性是完全独立的。这两个假设最后会使得从事研发活动的研究单位数目完全不一样。在第四部分中分析的模型相对简单。假设用以刻画研发技术特征的随机过程是一个泊松过程。这使分析得到了极大地简化。但是，在分析风险承担程度时，泊松过程的局限性也是很明显的。因此，在第五部分，我们假设了一个一般的扩散过程，并对需要回答的问题给出了部分答案。

之所以要将关注的焦点集中于这两个极端情形中，是因为它们能提供最为简单的研究背景。有了这一背景，我们就可以用一系列可供采取的研究策略来分析最终形成的市场均衡的性质的重要影响。在一个更一般的模型中，不同研究策略之间的相关关系本身就是内生的。因此，也许有人会问，市场经济中的研究策略是否充分的分散？或者各企业是否会不正当地模仿竞争对手的研究？下面将会看到，产品空间（product space，也就是产品多样化）上的垄断竞争与研究策略空间上的竞争非常相似。

当然，我们也意识到，我们在此所关注的纯粹的过程创新这一概念，本

4　阿罗（Arrow，1962）那篇很有影响的文章，并没有考虑研发活动中的竞争。阿罗的目标是对不同市场结构下，企业承担研发支出的激励进行比较。在该文所建立的正式模型中，虽然产品市场的结构是可以变化的，但是它通篇都假设只有一家企业开展研发活动，而且在研发活动这方面，该企业受进入壁垒的保护。因此，阿罗的正式模型就回避了研发活动中的竞争。

在本文的第三部分，我们将提供一个正式模型，用以刻画刚才在正文中所描述的环境。

身就是一个理想化的概念。我们所想到的创新通常都发生在商用飞机的制造上，有时也会发生在电子领域和摄像设备领域。就这些领域而言，忽略掉产品多样化问题也是有一定道理的。当然，一个更为完整的理论必须要同时考虑产品创新和过程创新，及其与产业结构的关系。

二、初步的结果

从本质上讲，下面的分析是局部均衡分析。第一，我们考虑一个给定商品的市场。若 Q 是总产出，则社会效用为 $u(Q)$。假设 $u'(Q)>0$ 且 $u''(Q)<0$。消费者的支付意愿（等于市场价格）假定为

$$p(Q) = u'(Q) \tag{1}$$

若 c 是由当前最有效的技术决定的单位产品的成本，则当前的社会净剩余为

$$s(c) = \max_{\{Q\}} u(Q) - cQ$$

我们将成本为 c 时，能最大化社会福利的 Q 的值记为 $Q_s(c)$。假设一项创新发生后，能使单位产品的成本从 c 降低到 c^*。如果该市场是由社会管理的，那么每期的社会净剩余增加量为

$$\pi_s = s(c^*) - s(c) = \int_c^{c^*} \frac{\partial s}{\partial c} dc = \int_{c^*}^{c} Q_s(c)\, dc \tag{2}$$

π_s 是从创新中获得的“社会利润”流。我们可以将

$$V_s = \pi_s / r$$

定义为社会利润的现值，其中 r 是（社会）贴现率。应该要注意到，在这一分析中，我们已经隐含地假设了不存在耐用资本品。引入耐用资本品会使分析变得更加复杂，但并不会从本质上改变结论。

我们也假设当前的创新不会被后来的创新所替代。很明显，这是一个不太现实的假设。再说一遍，可以去掉这一假设，但也会使分析变得更加复杂（Gilbert and Stiglitz，1979；Dasgupta and Stiglitz，1977）。

第二，假设该产业由一个完全垄断企业（即存在很高的进入壁垒）所控制。我们假设随着产出的增加，净边际收益会下降。当生产成本为 c 时，垄断企业每期的利润为

$$P(c) = \max_{Q} p(Q)Q - cQ$$

我们用 $Q_m(c)$ 来表示成本为 c 时的垄断产出。同时，当成本从 c 降低到 c^* 时，垄断企业的利润的增加量为

$$\pi_m = P(c^*) - P(c) \equiv \int_c^{c^*} \frac{\partial P}{\partial c} dc = \int_{c^*}^{c} Q_m(c)\,dc \tag{3}$$

我们再一次将垄断企业从创新中所获得的垄断利润的增加量的现值定义为

$$V_m = \pi_m / r$$

我们假定存在一个完善的资本市场，使得各企业所面对的利率等于社会贴现率。

第三，我们考虑这么一种情形：现有的技术可以自由获得，但新技术的发明人会获得一项技术专利权。使用这项技术，生产成本为 c。因此，在专利保护期限内，专利持有人每期所能获得的利润为（用下标"e"表示新进入的企业）

$$\pi_e(c) = \max_{\{Q\}}[p(Q)Q - cQ] \tag{4}$$

约束条件为

$$p(Q) \leqslant c$$

当对产品的需求弹性小于 1 时，专利持有人总会使价格达到最高水平，即 $p = c$。若对产品的需求弹性大于 1，且 c 足够小，则专利持有人会降低价格。

显然，$\pi_e(c) = 0$，因此

$$\pi_e(c) = \int_c^{c^*} \frac{\partial \pi_e}{\partial c} dc = \int_{c^*}^{c} Q_e(c)\,dc \tag{5}$$

创新的现值的大小，取决于专利保护期（T^*）的长短：

$$V_e = \frac{\pi_e}{r}(1 - e^{-rT^*}) \tag{6}$$

在第四种情形中，我们感兴趣的是，当现有的技术（c）由一家垄断企业控制，而专利由另外一家企业拥有的情况。这可能非常接近熊彼特所说的那种市场环境。在这种市场环境中，一个垄断企业会暂时被另一家企业所取代。[5]于是，在创新出现后，市场就成了一个双头垄断市场，而且赢得专利权的企业所获得的利润要取决于两个垄断企业之间的相互作用。

在后面的分析中，我们将指出，可以得到一些一般性的结论。这些结论独立于任何与两个垄断企业间相互作用的性质相关的假设。

在下一部分中，我们会将三种市场状况（只有一个垄断企业，在初始

5　然而，在本文中，我们没有给出一个完全动态的模型，以刻画每一种创新都被后面的创新所取代这一情形。

技术上有竞争；对专利的垄断；对初始技术的垄断，在新技术上可能会有不同的垄断企业）中的均衡与社会最优的研发模式进行比较。我们将关注开展研发活动的研究单位（企业）的数目和创新出现的期望速度。

由于在不同的市场环境中，（总）收益是不同的，所以结果也会有所不同。在我们后面的分析中，下面这组不等式是很有用的：

$$\pi_s > \pi_e \geqslant \pi_m \tag{7}$$

或者，在专利保护期限无穷大时，上式就等同于，

$$V_s > V_e \geqslant V_m \tag{7a}$$

图1描绘了这些变量之间的关系（到第三部分再介绍与双头垄断相关的基本不等式）。只要注意到，对于任何一个 c，都有

$$Q_s(c) > Q_e(c) \geqslant Q_m(c)$$

并运用式（2）、式（3）和式（4），立即就可以得到不等式（7）。

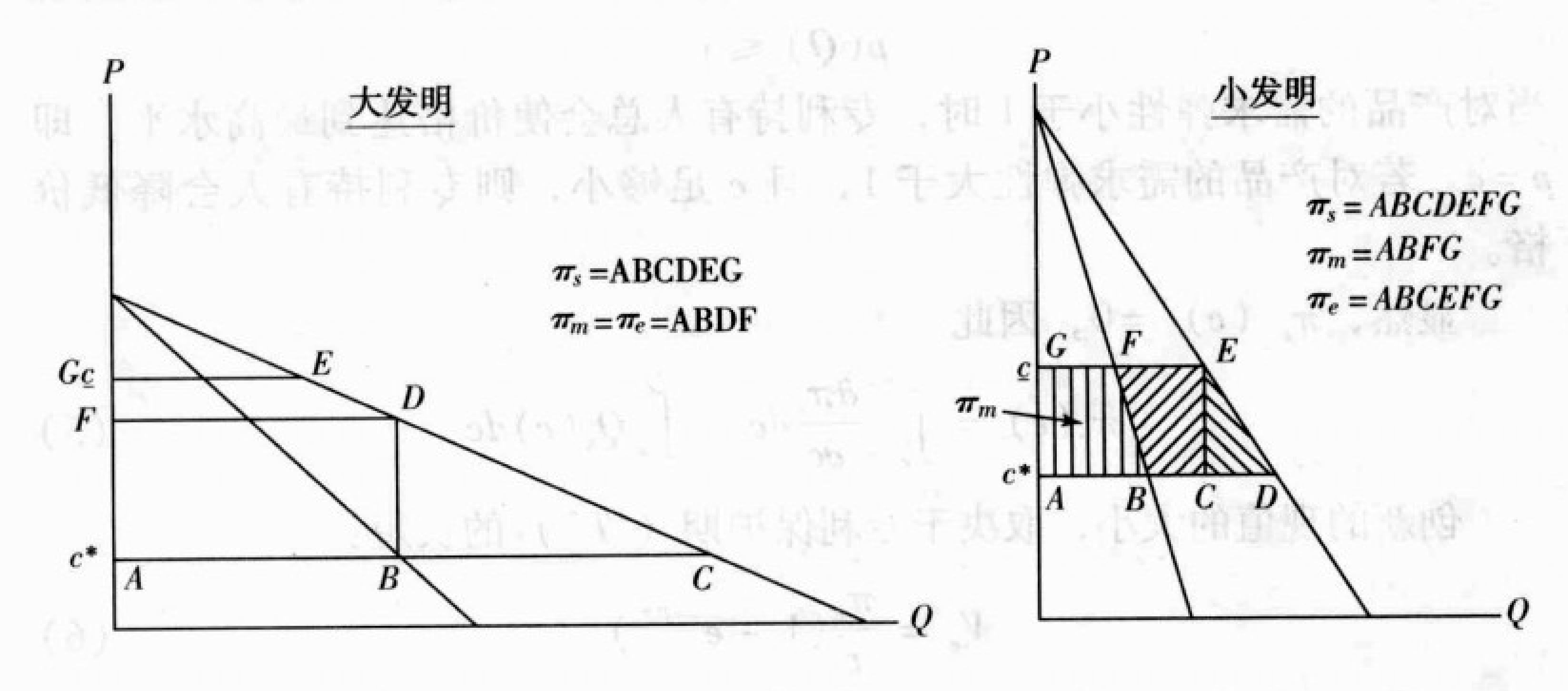


图1

社会收益（π_s）大于竞争收益（π_e），而竞争收益又大于垄断收益（π_m）这一事实表明，竞争性市场上的研究活动不足，而垄断市场上的研究甚至更少（此外，后一个结论似乎与传统的熊彼特式的分析相矛盾。这种分析认为，至少一定程度的垄断会对研发活动产生积极的影响）。然而，这些结论最多也是一些可疑的结论而已。要评价它们的正确性，就要将研发活动中的竞争模型化，同时还要区别产品市场的竞争和研发活动中的竞争。现在，我们就转向这些工作。

三、确定性条件下的创新激励

社会管理的市场。如果一个研究单位在 $t=0$ 时投资 x，就可以在 $T(x)$

时期解决研究中的所有问题，[6,7]即它会在 $T(x)$ 期“做出”发明。我们假设，对于所有$x \geqslant 0$都有 $T(x)>0$，并且投资越多，发现的日期就会越靠前。为了尽量避免角点解带来的麻烦，我们假设 $T(\infty)=0$ 和 $T(0)=\infty$。不过，很容易就可以去掉这些假设。此外，假设研发技术带来的边际收益是递减的，即$T''(x)>0$（见图2)。

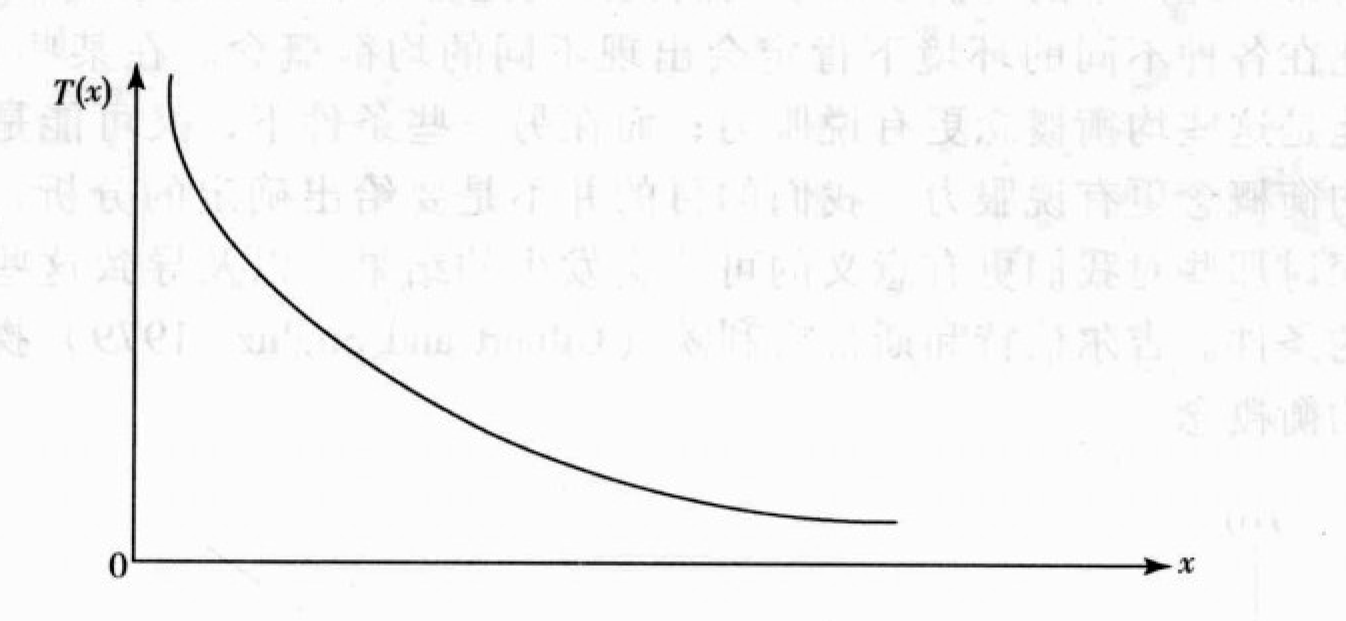


图 2　研发技术中的收益递减

首先考虑社会管理的市场。很明显，社会计划者的问题是：

$$\max_{x \geqslant 0}\left(V_s e^{-rT(x)}-x\right) \tag{8}$$

尽管我们假设研发技术上的收益是递减的，但是看式（8）就知道，这个最大值函数一般都不会是凹函数。这是过程创新的一般性特征。图 3 描绘了社会最优的研究期限。

完全垄断企业。现在，要将承担这种创新的社会激励，与在研发领域受到进入壁垒保护的垄断企业所承担的激励进行比较，就很容易了（在后面的小节中，我们将分析，当研发活动中不存在进入壁垒时，垄断企业承担研发支出的激励)。垄断企业的问题是，

$$\max_{x \geqslant 0}\left(V_m e^{-rT(x)}-x\right) \tag{9}$$

6　假定所有潜在的研究单位都面临相同的研发技术，只是为了便于解释而已。参见脚注 12。

7　另一方面，可以将 X 看成是，能在 T 期就实现创新的最优研究策略所要求的投入品的现值。当研究的结果是随机变量的时候，这两者就不相等了。这是因为投入会在创新出现的那一期被中止。因此，与产出一样，投入也是随机的。

对于应用程度很高的研究，这一模型也是适用的。比如，许多专家非常自信地宣称：新一代客机将会在 1981 年和 1982 年投入运行；波音 767、空中客车 A310 - 200 和洛克希德 L1011 - 400 的载客量和飞行里程也将大致相同；每一种机型的研发都将耗费 10 亿美元。在此我们想强调的是，专家们在宣称新机型交付日期时的自信。

令 x_m 为式（9）的解。由于 $V_s > V_m$，于是可以从式（8）和式（9）推出 $x_s > x_m$。[8] **因此，在研究部门受进入壁垒保护的垄断企业所承担的研发支出以及研究速度，都低于社会合意水平。**

竞争性市场：引论。 在此处及后面的讨论中，我们分析含有多家企业的市场经济中的均衡。由于我们关注的是各种不完全竞争的市场环境，因此在各种不同的环境下肯定会出现不同的均衡概念。在某些条件下，可能是这些均衡概念更有说服力；而在另一些条件下，又可能是其他一些均衡概念更有说服力。我们的目的并不是要给出确定的分析，而是要去探讨那些对我们更有意义的可能会发生的结果，以及导致这些结果出现的条件。吉尔伯特和斯蒂格利茨（Gilbert and Stiglitz，1979）探讨了其他均衡概念。

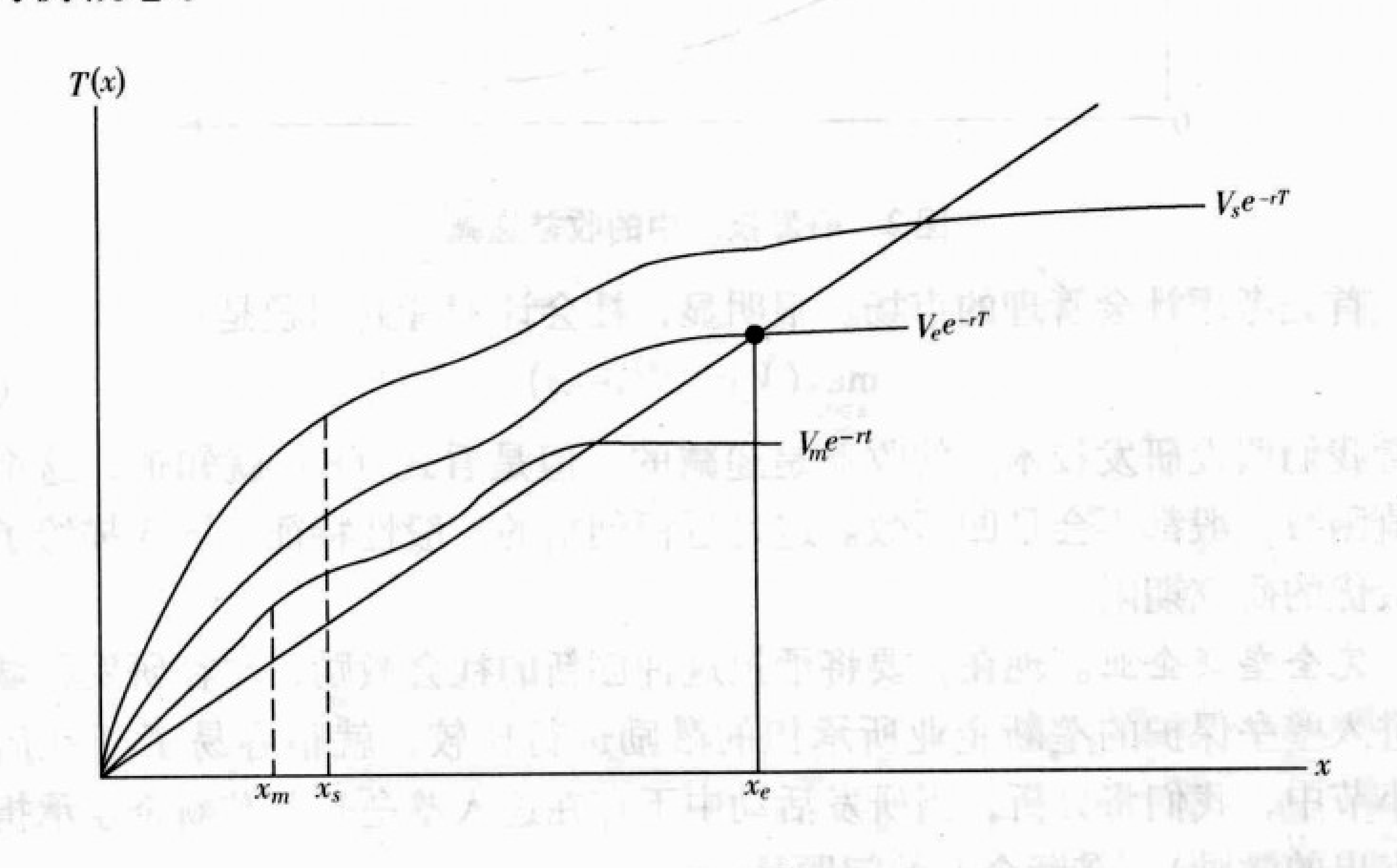


图 3　社会最优

在决定均衡的结构的各种因素中，下列因素是很重要的：

（1）各个企业可以采取的各种行动：在讨论纯策略纳什均衡和下面将会提到的另一个解时，我们假定企业只有一个决策变量 x。而在讨论动态策略时，我们假设每家企业都有一个研究计划，即在不同的日期有一系列的支出。后面各期的行动，依赖于竞争对手前期的行动。假定只有一个变量会极

8　读者可以证明，即使当式（8）和式（9）中的收益函数对 x 不是凹的，这一结论依然成立。

大地简化分析。对于那些初始投入很大的研究计划，或短期研究计划而言，这一假定是非常合适的。

（2）行动的先后次序，以及每家企业获得的与其他企业所采取行动有关的信息：在纯策略纳什均衡模型中，我们假设所有企业必须同时采取行动；在下面一小节，我们假设，一些潜在的进入者要等到垄断企业执行其研究计划后才采取行动，而且它们在采取行动之前，就可以观察到垄断企业的行动。

当存在研究计划（一系列的支出）时，每家企业可以获得的信息结构是非常重要的。原因在于，如果竞争对手以前各期的行动是可以观察到的，那么每家就可以预知，自己的行动对其竞争对手的行动的影响（显然，若竞争对手不能观察到它的行为，则它也不能预知这种影响）。

（3）对竞争对手行动的信念：一般的纳什均衡都会假定，每家企业都认为，竞争对手就自己的行动而做出的反应是不会改变的。下面我将探讨这种纳什均衡。我们的分析将表明，尽管在一些条件下存在混合策略纳什均衡，但是在纯策略中并没有纳什均衡。

在对另一个解和动态策略的讨论中，我们探讨的是一些完全不同的期望之中的结果。而且我们也认为这是一些更可能会出现的结果。可以将所有这些结果视为一些反应函数均衡。在下一小节中所探讨的特别的均衡，基本上就是大家非常熟悉的冯·斯塔尔伯格均衡的一个变体。与那些对称性更强的情形相比，对产业中有一个垄断企业的此类情形，这个均衡概念更有说服力。

纯策略中的纳什均衡。首先，我们进而考虑，没有任何一家企业能拥有 c 技术的专利权的这种情形。从而，我们假定，目前该商品的市场价格为 c。研发活动可以自由进入。假设大量企业的研发函数都是相同的（这个假设对下一小节的结论是很关键的。它意味着不存在边际内的企业，也没有利润（租金））。最先做出发明的企业会获得持续时间为 T^* 的专利权。如果是多个企业同时获得专利权，那么它们就在专利保护期限内共同分享市场（相互之间可能会进行古诺博弈）。[9]

在这一部分中，我们假设，各个企业只能在0期就投入水平 x 同时做出

9 由于我们假定只有一种发明，因此（在专利保护期限内）赢家通吃的假设最有道理。当然，在实践中，各企业都会围绕专利进行研究，所以最先做出发明的企业并不一定就最有优势。很明显地，可以将此模型拓展，从而引入这些特征。

唯一决策。为了使问题变得有意义，我们排除了没有任何企业开展研发活动这种可能的均衡结果。也就是说，假设存在一个正的 x 能使 $V_e e^{-rT(x)} - x > 0$（见图3）。

首先，在分析竞争性市场时，我们假设，所有企业（不管它是否承担研发支出）在考虑另一家企业的研发支出水平时，都会做出古诺推测。最先要注意到的是，只有一家企业会开展研发活动。为了确认这一结果，假设在一个潜在的均衡状态会有多家企业开展研发活动。显然，每家未能赢得专利权的企业都会有亏损，因此这就不能实现利润最大化。其原因在于，企业是可以关闭的。如果有两家企业展开研发竞争，那么在假定另外一家企业不会改变其研发支出的条件下，每家企业都会选择在边际上增加其研发支出，从而在边际上将把发明的日期提前，这样就可以确保另一家企业不能来分享专利权，因此会使其利润的现值增加一个离散的数量。于是，我们就可以得到一个结论：如果存在一个均衡，那么一定就只有一个企业。[10]

接着，我们注意到，在均衡状态（如果存在的话），该企业无法获得正的利润现值。原因在于，在自由进入条件下，一个会对该企业做出古诺推测的潜在的进入者，将会选择一个稍微高一点的研发支出水平，从而确保自己能独享专利权，并获得一个正的利润现值。这就表明，均衡可以由一家积极的企业来表征。该企业承担的研发支出为 x_e，而 x_e 就是下列零利润条件的最优解（见图3）。

$$V_e e^{-rT(x)} = x \tag{10}$$
[11]

但是，如果所有企业对所有其他企业的研发活动都进行古诺推测，上述结论就不成立了。原因在于，这家企业会认为，在所有其他企业都节约其研发支出的条件下，它就可以通过削减研发支出使其低于 x_e，来提高利润的现值。因此，我们就得到了以下结论：*如果所有企业对所有其他企业的研发支出水平进行古诺推测，那么纯策略中的纳什—古诺均衡就不存在*。当然，可能会存在一个混合策略均衡。在此，我们并不讨论这一问题（Gilbert and Stiglitz, 1979）。

另外一个解。上述结论意味着，积极的企业会对潜在的进入者进行古诺推测的思想，与熊彼特式的竞争思想并不一致。下面我们就需要引进进入威

10　我们已经通过假设排除了没有任何研发活动的均衡。

11　读者很容易就可以证明，式（10）有最优解，而且这个解就是唯一的均衡研发支出水平。

胁。可以通过假设积极的企业会影响潜在进入者的反应函数，来引进进入威胁。也就是说，在考虑潜在进入者的行为时，会进行冯·斯塔尔伯格推测。有了这个设定，一定会存在古诺—纳什均衡，而且均衡是唯一的。可以用承担研发支出水平 x_e 的企业来刻画这个均衡。x_e 是零利润条件式（10）的最优解。[12]

如果我们假设，潜在的进入者能依次进行研发投入（即它们不必同时进行决策），而且任何企业就 x 做出的决策都是不可更改的（更改决策需要付出巨额成本），以及任何一家企业的行动都能被所有其他企业观察到，那么这个解就是非常可行的。因此，由于我们假定大量企业都能获得相同的研发技术，那么就不会存在很多正的利润，而且有些企业也无法从这个机会中获利。此外，因为企业不可能认识不到这一点，所以如果它们开展研发活动，它们就会以一定的力度来进行研发。在该力度水平上，它们能阻止其他企业进入研发领域，并能防止其他企业对他们地位的削弱。[13]

根据式（7）和上述分析，就可以得到一个很直接的推论：竞争既会使研发支出高于社会最优的或垄断的配置，又会使研发支出低于社会最优的或垄断的配置。

有三个因素决定了 x_e，x_s 和 x_m 的相对大小：

（1）虽然社会最优的或垄断的配置要求边际收益等于边际成本，但是在为进入配置而展开竞争的条件下，平均收益要等于平均成本。由于平均收益大于边际收益，因此这会使研发支出水平过高。

（2）专利保护期限越长，x_e 就越大。显然，当 $T^* \to 0$ 时，有 $x_e \to 0$。

（3）比率 $V_s/V_e(V_e/V_m)$ 越大，$x_e < x_s(x_e > x_m)$ 的可能性就越大。比率 π_s/π_e 取决于需求函数的形状。为简化起见，我们关注弹性不变的需求函数。需求函数的形式为

$$p = Q^{-1/\epsilon} \tag{11}$$

12 如果不同企业面临不同的研究技术，分析也是很清楚的。在此仅举一例，假设存在一个潜在企业 i 的连续统（$0 \leqslant i \leqslant 1$），$T_i(x)$ 是研发支出为 x 的企业 i 做出发明的日期。假设对于所有的 $x \geqslant 0$ 和所有 $i > j$ 的 i 和 j，都有 $T_i(x) > T_j(x)$。显然，$i = 0$ 就是最有效的企业，而且只有这家企业会开展研发活动，尽管在均衡状态其利润现值为零。同时，似乎也有与此相关的经验证据。例如，麦道公司最近就决定调整公司计划以发展它的 DC-X-200 型飞机。这种机型将与波音 767 展开直接竞争。

13 显然，当只有一家企业垄断 c 技术时，这一分析就可以扩展到式（11）所确定的环境中去。有一个很自然的假设，即这个垄断企业处于先动地位，因此它就可以维持其垄断。

它是对下面这个效用函数求导后得到的

$$u = \frac{Q^{1-1/\epsilon}}{1-1/\epsilon} \tag{12}$$

因此，不失一般性地，令 $c=1$。同时，设

$$\pi_s = \frac{c^{*1-\epsilon}-1}{\epsilon-1} \tag{13a}$$

若 $\epsilon<1$ 或 $\epsilon>1$，且 c^* 不是非常小，则 $\pi_e=(1-c^*)$；若 $\epsilon>1$，且 c^* 很小，则

$$\pi_e = \frac{c^{*1-\epsilon}-1}{\epsilon^{\epsilon}(\epsilon-1)^{1-\epsilon}} \tag{13b}$$

以及只有当 $\epsilon>1$ 时，

$$\pi_m = \frac{c^{*1-\epsilon}-1}{\epsilon^{\epsilon}(\epsilon-1)^{1-\epsilon}} \tag{13c}$$

于是，假如 $\epsilon<1$ 或 $\epsilon>1$，且 c^* 不是非常小，则

$$\frac{\pi_e}{\pi_s} = \frac{(1-c^*)(\epsilon-1)}{c^{*1-\epsilon}-1} \tag{14}$$

而

$$\frac{\pi_e}{\pi_m} = \frac{(1-c^*)}{c^{*1-\epsilon}-1}(\epsilon-1)\left(\frac{\epsilon}{\epsilon-1}\right)^{\epsilon} \tag{15}$$

因此，如果 $\epsilon<1$，则对于任何给定的创新，π_e/π_s 会随着 ϵ 的增加而减少。然而，对于一个较大的创新（c^* 较小），π_e/π_s 会较小。

于是，在其他条件不变时，弹性更大的需求曲线就更有可能与竞争性市场上的过度研究有关，虽然这时也有一些与“重大创新”有关的不适当的研究。市场经济中过多的研发支出并不是来自于重复性的研究努力，而是来源于竞争压力。竞争会迫使均衡状态中的企业在社会合意的日期之前就做出创新。

如果我们以一种特别的形式引入发明日期上的不确定性，即所有企业都只能采取一种研究策略（所以所有企业都遵循相同的决策树），而且所有不确定性都是完全相关的，那么有自由进入的古诺—纳什均衡的上述特征（即只有一家企业开展研发活动，而且它所获得的（期望）利润为零）都会保留下来。这就意味着，给定其他企业的研发速度，一家特定的企业就可以保证：通过选择一个充分快的研究速度，就可以最先做出发明。当然，我们不能确定，在哪个时间它的研究部门会完成所有研究任务。当不确定性为这种形式时，我们可以直接得知：在均衡状态，只有一个企业会承担研发支

出，而且如果各个企业是风险中性的，那么该企业的利润现值的期望值就等于零。

上述结论非常重要。因为它告诉我们，开展特定研发活动的企业的实际数目，并不是竞争程度的一个指标。对于我们的模型而言，尽管只有一个企业从事研发活动，但是竞争也非常激烈。该企业的研发支出水平要足够的高，这样才能阻止其他企业进入研发领域。

动态策略。前面的讨论让每个企业只执行一个决策变量 x。然而，正如我们先前所指出的那样，绝大多数研究项目都会持续多期。在0期，企业并不能执行整个研究项目。这就意味着，如果它的竞争对手的行动是可以观察的，那么它就可以预知其行动对竞争对手行动的影响。

假定各个企业依次“采取行动”。并假设随机选出一个企业，由它最先采取行动。此外，我们假设，如果一个企业是垄断企业，那么在第一期它只会花费少量资源用于开展研发活动。

更精确一点来说，如果 $x(t)$ 为最先行动的企业在 t 期的研究力度，$c(x)$是研究计划 $\{x(t)\}$（以现值衡量）的成本，同时我们将发现的日期视为整个研究计划函数，那么企业最优的研究支出就是能最大化下式的 $\{x\}$

$$V_e e^{-rT(x)} - c(x) \tag{16}$$

我们假设最优解中包括在第一期严格为正的 x。现在假设最先行动的企业在第一期支出一个严格为正的 x，并宣布其他任何企业想参与竞争，那么它就会选择一个能确保其获得专利权的研究策略（如果竞争对手只实施能获得非负利润的研究策略）。显然，若所有企业都面临同一个创新成本函数，则任何能使潜在竞争对手实现收支平衡的政策，也能使最先行动的企业实现收支平衡。况且，对它而言这种策略确实是最优的策略。其原因在于，实施这种策略能完全阻止其他企业进入研发领域。但是，假如其他企业都不想参与竞争，那么最先行动的企业自然就会选择垄断企业的政策。

令 $T(x_{mc})$（表示“垄断—竞争”）为式（16）的解。比较式（8）和式（16），就可以发现，只要 T^* 足够大，就有 $T(x_{mc}) > T(x_s)$。在这些条件下，我们就可以说，在竞争性市场中扩展创新活动的激励会低于社会合意的水平（Arrow，1962，p. 152）。我们已经说明了，如果各企业不必非得开展研发活动，那么即使研发领域能自由进入，一家给定的企业也能够有效地阻止其他企业进入，并可以像垄断企业那样开展其研发

活动。[14]

我们也可以将这个均衡视为纳什均衡：给定最先行动的企业所宣布的策略，任何企业进入研发领域都得不到回报；而且，若没有其他企业进入，则最先行动的企业最优的选择就是执行垄断企业的研究计划。[15]

双头垄断与持续地垄断。在前面的讨论中，我们已经分析了两种市场情形。在第一种情形中，一个企业控制目前的技术，并且只有该企业才可以（能够）进行研究。在第二种情形中，当前的技术可以自由获取，而且有许多企业都能开展研究活动。当然，研究竞赛的“优胜者”将（可能暂时）垄断新技术。

这一部分关注第三种市场环境。在这一种市场中，当前的技术（c）由一个专利持有人控制，但是在研发领域仍可以自由竞争。

首先假设（由于法律或技术上的原因）当前的垄断企业不能进行研究。正如我们在第二部分所注意到的那样，当发明出现后，市场就成了双头垄断市场。令 π_e^d 和 π_m^d 分别表示发明出现后进入者的利润流和垄断企业的利润流，V_e^d 和 V_m^d 分别表示相应的利润流（贴现到发明出现时）的现值。在双头垄断市场上，成功进入者的利润，通常都高于存在大量竞争者时所获得的利润（若可以用一个定量的纳什均衡来描述进入后的市场，则这一点将会变得更加清楚）。

因此，通常会有

---

14 例如，考虑一个两阶段研究计划。令 $S\{\hat{x}(t)\}$ 为零利润研究策略集合，也就是说，

$$V_e e^{-rT(\hat{x}(1),\hat{x}(2))}=c(\hat{x}(1),\hat{x}(2))$$

令 x_m（1）为垄断企业在第一期用于研发的支出。令 $\hat{x}_m(2)$ 为 $x(2)$ 的值。给定 $x_m(1)$，并将其视为沉没成本，则 $\hat{x}_m(2)$ 恰好能使最先行动的企业收支平衡。也就是说，

$$V_e e^{-r\langle T(x_m(1),\tilde{x}_m(2))-1\rangle}=\tilde{x}_m(2)$$

我们假设，$T(x_m(1),\tilde{x}_m(2))<T(\hat{x}(1),\hat{x}(2))$ 对于所有 $\hat{x}(1)$，$\hat{x}(2)\in S$（即能达到收支平衡的研发支出）都成立。显然，这不一定就是正文中所说的那种情形。如果它不是那种情形，那么最先行动的企业要有效地阻止其他企业进入，就得有 $x(1)>x_m(1)$。但是，对于阻止进入策略而言，将依然会有正的利润。

然而，如果“领先者”无法观察到竞争对手采取的行动，那么在竞争对手赶上它之前，“领先者”可能不会采取行动以“抵御”竞争者。吉尔伯特和斯蒂格利茨（Gilbert and Stiglitz，1979）分析了这一情形中的均衡的性质。

15 确实，当进入者的策略要求其研究计划能满足下面这一条件时，它甚至是一个完全的纳什均衡

$$V_e e^{-rT(x)}\geqslant c(x)$$

$$V_e^d > V_e \tag{17}$$

然而，研发活动上的自由进入体现的就是零利润：

$$V_e^d e^{-rT(x_d)} = x_d \tag{18}$$

比较式（10）和式（18），并运用式（17），我们立刻就可以得到

$$x_d > x_e$$

也就是说，当市场最初是由一个垄断企业所控制时，其研发支出水平要高于竞争性市场中的研发支出水平（在这一意义上，垄断促进了研发）。

紧接着，由于 x_e 也许会大于 x_s，因此可能就会有过多的资源用于研发活动：实际上，在为成为第一个创新者的竞赛中，所有潜在的垄断利润都消散了。

现在，我们转而考虑垄断企业可以进行研发的情形。我们现在指出，**现有的垄断企业将会通过花费足够多的资源用于研发，来阻止潜在的竞争者进入研发领域，从而确保它自己能获得新的专利权。**[16]不过，垄断企业也获得了超额利润。此外，垄断企业也许会以快于社会最优水平的速度来进行研究。为了看清楚这一点，将发明出现后垄断企业（最大的）利润流表示为

$$\pi_m(c^*) = \max_Q p(Q)Q - c^*Q$$

并且令 $V_m(c^*)$ 为相应的利润流（贴现到发明出现时）的现值。由于双头垄断市场上的两个企业一般都不会开展共同利润最大化的活动，所以就有

$$V_m(c^*) > V_e^d + V_m^d \tag{19}$$

因为竞争者不会选择一个高于 x_d 的研发支出水平（这是由式（18）确定的），所以垄断企业只需要使其研发支出略高于 x_d，就可以保证它自己能获得新的专利权。若垄断企业这么做，则利润的现值为

$$V_m(c^*)e^{-rT} + \frac{\pi_m(c)(1-e^{-rT})}{r} - x > (V_e^d + V_m^d)e^{-rT} +$$

$$\frac{(1-e^{-rT})\pi_m(c)}{r} - x = V_m^d e^{-rT} + \frac{\pi_m(c)(1-e^{-rT})}{r}$$

这也是垄断企业不采取先占行动时的利润的现值（可以由式（19）得到第

16 达斯古普塔、吉尔伯特和斯蒂格利茨（Dasgupta，Gilbert and Stiglitz，1979）在自然资源的利用时，就注意到了这一点。可以从吉尔波特和纽伯利（Gilbert and Newbery，1979）的文章中看到更为一般的版本。对该模型的综述，参见 Salop（1979）。

一个不等式，将式（18）代入，就可以得到后面的等式）。[17]因此，垄断企业总会通过支付超过 x_d 的研发成本（可能只是稍微超出一点点）来阻止竞争者的进入。

如前所述，假设垄断企业在研发领域受进入壁垒保护，而使其利润最大化的研发支出为 x_m（即 x_m 为问题（9）的解）。我们已经证明了，若 $x_m \geqslant x_d$，则研究领域中自由进入这一事实，并不会使先前那些与创新激励有关的结论有任何改变。先前那些结论是在垄断企业在研发领域受进入壁垒保护的条件下得到的。不过，如果 $x_m < x_d$，那么垄断企业面临着研发竞争这一事实就会对结果产生重大影响。垄断企业将会通过支出超过 x_d 的研发成本（可能只是稍微多一点点）来阻止其他企业进入研发领域。在这一情形中，由于垄断企业必须要支付高于其理想水平的研发成本，所以它所能获得的利润流的现值就会下降。但是，不管怎么说，它还是能获得超额利润。最后，我们将会看到，如果社会最优的研发支出水平 x_s 低于 x_d，那么面临研发竞争的垄断企业就会开展过多的研究。

从直观上讲，与潜在的竞争者相比，垄断企业总会占有优势的原因是很清楚的。如果另外一家企业赢得专利权，那么产业结构就变成双头垄断结构。当存在自由进入时，这个后来进入的垄断企业就会意识到，它要想与现有的垄断企业分享市场份额，它就要支付足够的研发支出，直到它所获得的利润流的现值为零（即它会使 $x = x_d$）。然而，现有的垄断企业总是通过可以使其研发支出稍高于其他潜在的竞争者能获利时的研发支出水平，来维持其现有的垄断地位。关键是要注意到，这样做总是符合垄断企业的利益的。其原因在于，通过维持垄断地位，垄断企业所能获得的利润流就会高于它在双头垄断结构下获得的利润总额。当然，如果在研发活动中现有的垄断企业的效率要高于其竞争者，那么这一观点就会得到加强。这些结论的含义是，即使在研发活动中存在竞争，一个垄断企业仍然很有可能会继续保持垄断状态。垄断企业受到潜在竞争者威胁的这一事实，最多也就是促使它支付更多的研发成本而已。但是，对于在我们所分析的模型中，该产业仍然会是一个垄断产业。

我们在此所使用的解的概念，与前面所概述的解相类似。在这个解中，

17　这一观点比我们在此所给出的分析更具有一般性。例如，我们可以引入耐用资本品（此时，最大化的利润的现值就不像正文中所讲的那么简单了）和不确定性（Gilbert and Newbery，1979）。正如我们在下面将会看到的那样，所有进入者只能获得零利润这一假设是非常关键的。

垄断企业像一个冯·斯塔尔伯格领先者那样行事，并承诺其研究投入水平为 x，其他进入研发领域中的企业获得的利润为零。在这种背景下，使用冯·斯塔尔伯格解可能是很有说服力的。[18]但是，那样的"承诺"和"零利润"假设并没有太大的说服力。当去掉这些假设时，这个解的现值就会发生很大的变化，例如，一方面先占（preemption）可能不再是合意的结果（Gilbert and Stiglitz, 1979），[19]另一方面，先占也许并不能使其研究计划显著不同于，一个没有竞争威胁的纯垄断企业的研究计划（如前所概述的那样）。

四、不确定性下的研究速度

引论。[20]在上一节中，我们得到的一个令人惊讶的结论是，即使存在专利制度且研究部门也可以自由进入，在没有不确定性的条件下，最多只有一家企业会开展研究。此外，我们注意到，如果各企业都要支付研发成本，同时如果所有企业都能自由获取现有技术，那么创新的速度会快到使利润的现值恰好为零，从而阻止更多的企业进入。假如在研究成功的日期上存在不确定性，但是如果所有企业面临的不确定性在前面所讲的意义上是完全相关的，那么这个结论仍然会成立。但是，若这两个假设不成立，则情况就会不一样。由于没有一家企业肯定会赢得专利权，因此进入和研究速度就共同使得期望利润为零。

现在，我们假设，发现（discovery）的日期是随机的。更精确地说，如果一家企业在 $t=0$ 时付出的研发成本为 x，那么它在 t 时点或在此之前做出发明的概率为 $1-e^{-\lambda(x_i)t}$。也就是说，在时间间隔（t，$t+\Delta t$）内，做出发明的概率恰好等于 $\lambda(x_i)\ \Delta t$。我们假设，$\lambda(x_i)$ 的特征可以表示为，在最初的一个范围内，收益是递增的，接着收益就会递减。图 4 描绘了这个特征。

---

18 吉尔伯特和斯蒂格利茨（Gilbert and Stiglitz, 1979）研究了此问题的（包含混合策略的）纳什均衡的结构。

19 这个观点是很简单的：假设研究结果是不确定的，并且不同企业可以采取许多不同的研究计划。此外，假设某个企业相对于其他进入者而言具有比较优势（但对垄断企业它并不一定就有比较优势）。所以，即使存在进入威胁，它的利润（实际租金）也为正。因此，为了阻止这种竞争者进入研发领域，现有的垄断企业就得大幅提高其研究支出。

20 本部分所阐述的模型最初出现在斯蒂格利茨（Stiglitz, 1971）的文章中。劳瑞（Loury, 1979）独立发展了这一模型。劳瑞关注的是这么一种情形，即 $V_s=V_e$，且各企业都会考虑自身的行为对发现日期的总分布的影响。我们主要考虑 $V_s>V_e$ 这种正常的情形。

为了简化分析，假设所有研究单位采取不同的研究策略，于是不同单位所面临的不确定性是相互独立的。但是，我们假设它们一样有效。因此，如果对第 i 个单位的投资额为 x_i，并且总共有 n 个这样的单位，那么在 t 时刻（从 $t=0$ 时算起）做出发明的概率密度函数由下列指数式给定

$$\sum_{i=1}^{n}\lambda(x_i)\exp[-\sum_{i=1}^{n}\lambda(x_i)]t$$

发现的期望日期为 $[\sum_{j=1}^{n}\lambda(x_j)]^{-1}$。

在下面两个小节中，我们将会分析研发由社会统一管理的情形，以及开展研发活动的是受到进入壁垒保护的垄断企业的情形。然后，我们再研究竞争性市场这种情形，并将三种情形下的结论进行比较。

社会管理的经济。由于研究单位是相同的，所以在最优状态我们将它们都视为一样的。与前面一样，令 V_s 为从发明做出时起获得的净社会利润流的现值（贴现到发现的日期）。目标就是使净社会剩余的现值的期望值最大化。因此，规划问题就是选择 n 和 x，使得下式最大化

$$V_s\int_0^{\infty}n\lambda(x)e^{-(n\lambda(x)+r)t}dt-nx \tag{20}$$

假设 n 是一个连续变量。[21]最优地选择 n 就得到一阶条件

$$\frac{V_s\lambda(x)r}{(n\lambda(x)+r)^2}=x \tag{21}$$

最优地选择 x 会得到[22]

$$\frac{V_s\lambda'(x)r}{(n\lambda(x)+r)^2}=1 \tag{22}$$

式（21）和式（22）就决定了 n 和 x。现在注意到，可以从上述两式中直接推出：在最优状态，有

$$\lambda'(x)=\frac{\lambda(x)}{x} \tag{23}$$

令 x^* 为式（23）的解。因此，对每个实验室的最优投资水平（见图4），与企业的数目无关。设实验室的最佳数目为 n_s。如果我们将 x 的值 x^* 代入式（21），那么 n_s 就是式（21）的解。

完全垄断市场。对于完全垄断企业（受到进入壁垒保护的企业）而言，

---

21 假如每个研究单位的有效规模 x^* 都"很小"，就可以证明这一假设是合理的。

22 只要 $\lambda'(0)>r^2/\pi_s$（我们将假设这一不等式成立），式（21）和式（22）就成立。

分析过程是一样的，只要将式（20）至式（23）中的 V_s 替换成 V_m 即可。因此，垄断企业能以有效的水平运营每个实验室。但是，实验室的数目不是最优的。如果 n_m 是能使垄断企业的期望利润最大化的研究单位数目，那么它就是下式的解

$$\frac{V_m\lambda'(x^*)r}{(n\lambda(x^*)+r)^2}=1 \tag{24}$$

由于 $V_s>V_m$，因此，比较式（22）和式（24）就可以发现 $n_s>n_m$。因此，垄断企业会推迟创新。创新出现的日期的期望值会更大，即预期要在更远的将来方可实现创新。这一结论与我们在第三部分中得到的结论类似。

完全竞争市场。现在，我们转而研究竞争性企业承担研发支出这一情形。各企业（实验室）是相同的，且相互独立工作。令企业 i 在 $t=0$ 时投资 x_i 用于研发，n 为进行研发投资的企业数目。企业 i 知道这个数目和选择 x_j（$j\neq i$）。于是，（从 $t=0$ 时看）企业 i 能在 t 时刻做出发明并赢得专利权的概率密度为

$$\lambda(x_i)e^{-\sum_{j=1}^{n}\lambda(x_j)t} \tag{25}$$

与前文一样，我们假设自由进入。因此，n 是内生的。与以前一样，我们令 V_e 为第一个发现者从发现之日起获得的利润流的现值（假设专利保护期限 T^* 为正）。假设各企业都是风险中性的。于是，企业 i 就希望选择 x_i 让下式最大化

$$V_e\int_0^{\infty}\lambda(x_i)e^{-(\sum_{j=1}^{n}\lambda(x_j)+r)t}dt-x_i \tag{26}$$

接下来，可以从两个方向展开分析。一个方向是，假设各企业都是研发精密的计算器，它们会考虑 x_i 对 $\sum_{j=1}^{n}\lambda(x_j)$ 的影响。另外一个方向是，假设每家企业都会忽略掉 x_i 对 $\sum_{j=1}^{n}\lambda(x_j)$ 的影响。这就是说，每家企业都将发明的期望日期当作给定的，并且只通过调整 x_i 来改变最先做出发明并获得专利权的概率。如果企业数目“很多”（这就是我们将要假设的情形），那么这个方向的分析就是可行的。因此，第 i 个企业选择 x_i 就得到[23]

$$V_e\frac{\lambda'(x_i)}{(\sum_{j=1}^{n}\lambda(x_j)+r)}=1 \tag{27}$$

23 在推导式（27）的过程中，我们假设 $\lambda'(0)>r/V_e$。

由于我们假设所有企业都是一样的，所以就能得到对称的古诺—纳什均衡。因此，式（27）就转化为

$$V_e \frac{\lambda'(x)}{(n\lambda(x)+r)} = 1 \tag{28}$$

此外，由于存在自由进入，所以对于每家企业而言，利润的期望现值都为零。于是，由式（26）我们可以得到

$$V_e \frac{\lambda(x)}{n\lambda(x)+r} = x \tag{29}$$

式（28）和式（29）就表示自由进入的古诺—纳什均衡，而且 n 和 x 都可以从这两个式子中解出来。

式（28）和式（29）有两个直接的含义。首先，

$$\lambda'(x) = \frac{\lambda(x)}{x}$$

因此，在这个均衡中，每个企业都能有效地运行各自的实验室，即 $x = x^*$（图4）。

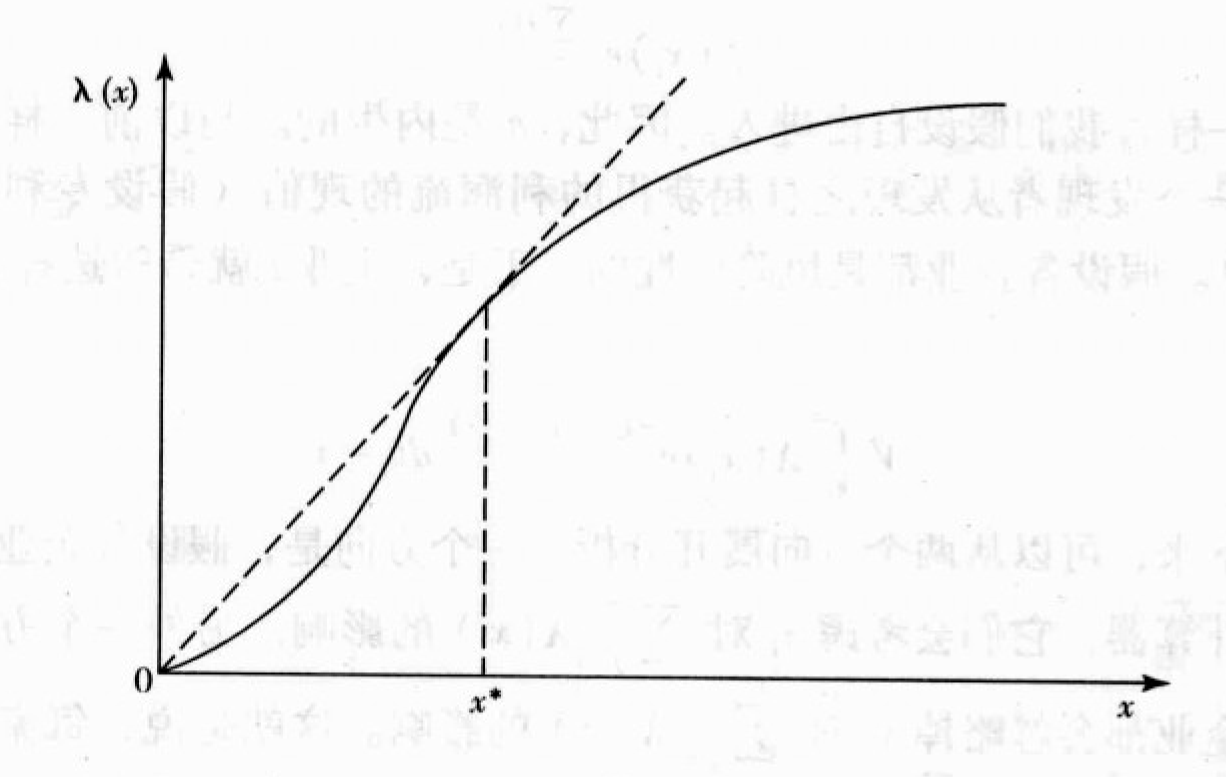


图4

由于在我们所研究的三种制度环境中，实验室都能得到有效的运行，所以为了评价不同经济环境中的研究速度，我们只需比较研究单位数目的多少。特别地，模型预示对于我们所研究的三种市场结构而言，发明的期望日期 $1/n\lambda$ 只是与研发总支出 nx 成反比例，而且比例常数为 $1/\lambda'(x^*)$。研发

支出总额 nx 与发明的期望日期之积，独立于市场结构。[24]

此外，如果任何研究都是在社会最优的状态中展开的，那么社会收益必定为正

$$\frac{V_s n\lambda}{n\lambda + r} > nx$$

在自由进入条件下，当 $V_s \approx V_e$ 时，就会有过多的企业从事研究。假如专利权的获得者从来自于研究的社会总收益中获得利益，那么在自由进入条件下，平均收益必须等于平均成本。于是，所有潜在的社会收益都会以过度进入的方式消耗掉（这不过是共有资源问题的又一个例子而已。每个进入者要么是获得所有利益，要么什么都得不到，但是它的期望收益只有专利权价值的 $1/n$）。[25]边际收益与平均收益的比恰好是 $r/(n\lambda + r) < 1$。

然而，正如我们所指出的那样，一般都有 $V_s > V_e$，于是，研究到底是太多了还是太少了，这一点尚很模糊。不过，至少在本文的简单模型中，我们可以确定一些条件。在这些条件下，就能判断出研究是太多了还是太少了。

由式（21）和式（29），可以得到

$$\frac{n_c\lambda(x^*) + r}{n_s\lambda(x^*) + r} = \frac{V_e(n_s\lambda(x^*) + r)}{V_s r} \tag{30}$$

在式（30）的右边运用式（21），我们就可以将式（30）化为

$$\frac{n_c\lambda(x^*) + r}{n_s\lambda(x^*) + r} = \frac{V_e\sqrt{\lambda(x^*)}}{\sqrt{V_s}\sqrt{x^*}\sqrt{r}} \tag{31}$$

这样，就可以得到，

$$\text{当 } V_e\sqrt{\frac{\lambda(x^*)}{x^*}} > \sqrt{V_s r} \text{ 时}, n_c > n_s; \text{而当 } V_e\sqrt{\frac{\lambda(x^*)}{x^*}} < \sqrt{V_s r} \text{ 时}, n_c < n_s。 \tag{32}$$

式（32）所表示的关系是非常有用的。一方面，我们可以从中计算出

---

24 在最近的一项研究中，施沃茨曼（Schwartzman，1977）指出，在制药业中，每一种新化学药品的平均研发成本大约为每年 1 700 万美元，平均而言，发明期限大约为 4 ~ 5 年。我们取最大值，并且不考虑折旧，这就意味着，每一种新化学药品的研发投资总额大约为 8 500 万美元。在我们的模型中，这就意味着，

$$\bar{x}/\lambda(\bar{x}) = n_c\,\bar{x}E(T(\bar{x})) = 85 \times 5 \times 10^6$$

美元年。而且模型的分析表明，这与市场结构无关。

25 斯蒂格利茨（Stiglitz，1971）和巴泽尔（Barzel，1968）最先指出了这一点。劳瑞的分析就限制在这类特殊的情形之中。

最优的专利保护期限为

$$-\ln\left\{1-r\frac{\sqrt{\pi_s/\lambda'(x^*)}}{\pi_e}\right\}\Big/r \tag{33}$$

假如

$$r\sqrt{\pi_s/\lambda'(x^*)}<\pi_e \tag{34}$$

（如果式（34）不成立，那么对于一项保护期限为无穷大的专利权而言，市场上的研究就是不充分的）。

另一方面，若给定模型的具体参数，我们就可以很容易计算出（有具体的专利保护期限的）市场上的研究是否过多了。为了阐明这一点，我们在此关注不变弹性的需求曲线，而且弹性小于1（所以赢得专利权的企业会把价格定在最高水平）。同时，假设专利保护期限无穷大。回忆一下第三部分的内容，若我们令 $c=1$，则有

$$\pi_s=\frac{1-c^{*1-\epsilon}}{1-\epsilon}$$

和

$$\pi_e=1-c^*$$

将上述两式代入式（34），并对 c 微分，我们就可以看到（见图5），对于小的创新而言，市场总会提供不充分的研究（将这一结论与第三部分中相应的结论进行比较）。另一方面，当 $c^*=0$ 时，则 $\pi_s=1/1-\epsilon$，且 $\pi_e=1$。因此，如果 T^e 表示发现的期望日期（$1/\lambda n$），且 R 代表总的研发支出（nx），并假设对于所有充分长的专利保护期限，都有

$$r<\sqrt{\frac{1-\epsilon}{R\cdot T^e}}$$

那么市场就会付出过多的研发成本（回忆一下，RT^e 是一个独立于市场结构的常数）。于是，假如利率充分低、需求弹性充分小，那么市场上就会有过多的研究单位从事开发专利保护期限很长的技术。应该要注意到，社会资源的浪费并不是来源于重复研究（与早期 Dasgupta and Stiglitz, 1980 的研究结论一致）。

在所有企业都采用同一种研究策略的情形中，除了一家企业之外，其他企业的所有研发支付都浪费了。而在我们所研究的情形中，由于每家企业都有一个研究策略，而且各自的研究策略是互不相关的，因此就能从每一个额外的研究单位中获得正的社会收益。在此处，社会资源的浪费是非常隐蔽的：它是由在某些特定的条件下边际社会收益会小于边际私人收益这一事实

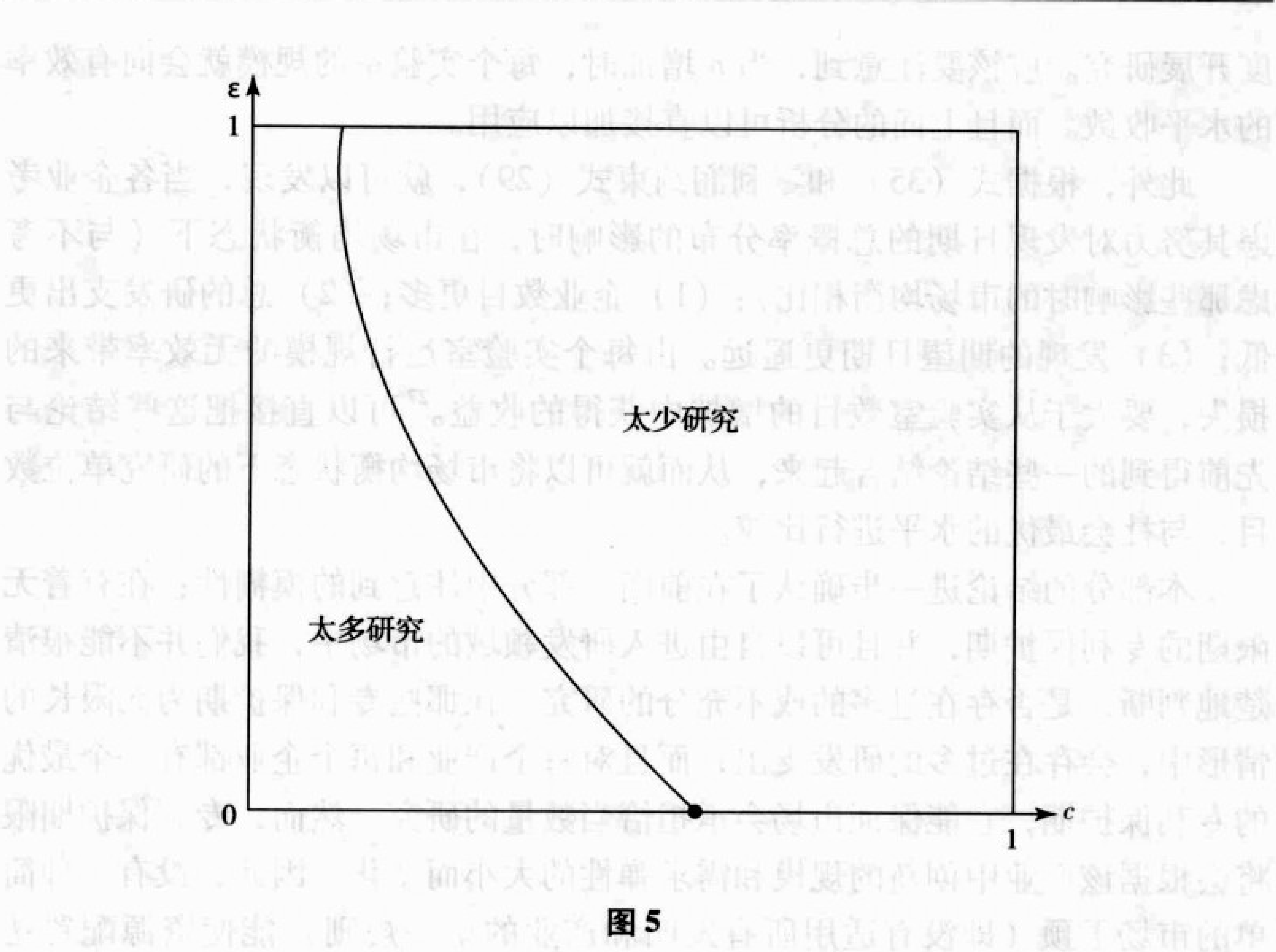


图5

引起的。在这里，边际私人收益是指赢得专利权的概率增加的那一部分。

应该要注意到，在其他条件不变时，需求弹性更低的产业更有可能出现过多的研发支出。

运用本文中的简单模型，我们还可以进行简单的比较静态分析。我们知道，在其他条件不变时，如果在一个产业中，创新成功后能获得更大的收益，那么该产业中就会有大量企业想去开发成本节约型技术。[26]

假如企业能认识到它的研发努力对发现日期的总概率分布的影响，我们就可以得到

$$\frac{\lambda' V_e}{n\lambda + r}\left\{1 - \left[\lambda\int_0^\infty te^{-(n\lambda+r)t}dt\right](n\lambda + r)\right\} = 1$$

或

$$\frac{\lambda' V_e}{n\lambda + r}\left[1 - \frac{\lambda}{n\lambda + r}\right] = 1 \tag{35}$$

在此情形中，直接就可以得知，每个实验室都将会以低于社会最优水平的速

26 在新近的一系列研究中，雅各·施穆克勒（Jacob Schmookler, 1962）强调了产品需求的增长对产业创新活动的影响。正文中前述结论从理论上确认了施穆克勒所关注的那种关系。

度开展研究。应该要注意到，当 n 增加时，每个实验室的规模就会向有效率的水平收敛。而且上面的分析可以直接加以应用。

此外，根据式（35）和零利润约束式（29），就可以发现，当各企业考虑其努力对发现日期的总概率分布的影响时，在市场均衡状态下（与不考虑那些影响时的市场均衡相比）：（1）企业数目更多；（2）总的研发支出更低；（3）发现的期望日期更遥远。由每个实验室运行规模的无效率带来的损失，要大于从实验室数目的增加中获得的收益。[27]可以直接把这些结论与先前得到的一些结论结合起来，从而就可以将市场均衡状态下的研究单位数目，与社会最优的水平进行比较。

本部分的结论进一步确认了在前面一部分中注意到的模糊性：在有着无限期的专利保护期，并且可以自由进入研发领域的市场中，我们并不能很清楚地判断，是否存在过多的或不充分的研究。在那些专利保护期为无限长的情形中，会存在过多的研发支出；而且对每个产业和每个企业都有一个最优的专利保护期，它能保证市场会承担恰当数量的研究。然而，专利保护期限将会根据该产业中创新的规模和需求弹性的大小而变化。因此，没有一种简单的市场干预（即没有适用所有发明和产业的单一规则）能使资源配置达到社会最优水平。此外，如果各企业考虑其研发努力对发现日期的总概率分布的影响（从理论上讲，这个假设是否合理，要看发现的日期是否依赖于研究者的数量），那么，即使是对特定产业进行特定干预，最优的专利保护期限也不能保证一定能达到社会最优。其原因在于，每个实验室都可能会在一个较小的规模上无效率地运行。然而，从理论上讲，可以通过采取对特许

27　把零利润条件代入式（35），我们就可以得到，

$$\frac{\lambda' x}{\lambda}\left(1-\frac{x}{V_e}\right)=1$$

从这个式子中，我们可以求出 x_e。由于零利润约束意味着，

$$n=\frac{V_e}{x}-\frac{r}{\lambda}$$

而至少对于较大的 n，就有

$$\frac{dn}{dx}=-\frac{V_e}{x^2}+\frac{r\lambda'}{\lambda^2}\leqslant 0$$

此外，还可以得到，

$$\frac{dnx}{dx}=\frac{r}{\lambda}\left(\frac{\lambda'}{\lambda}x-1\right)>0$$

$$\frac{dn\lambda}{dx}=\frac{V_e}{x}\left(\lambda'-\frac{\lambda}{x}\right)>0$$

经营权征税和其他税收工具，并确定最优专利保护期限，来达到最优状态。

五、不确定性下的研究速度——一般化的模型

在前一部分中，对研究过程所做的随机设定，有一个非常不好的特点：在不同时变动到达的平均时间的条件下，泊松过程不允许变动研究过程的风险。例如，我们想知道，当一系列研究项目的到达时间保持不变时，市场在判断风险更高的项目是否会存在偏差。

在本部分中，我们阐述一个简单的模型，以便能处理上述问题和其他一些问题。我们将 0 期的知识的“普通”状态用一个值（比如 A_0）来表示。可以将 A_0 视为，要使一个想法在商业上变得可行而必须克服的障碍的大小。我们进行研究的目的是为了解决大量小问题。每解决一个小问题，都会使 A 的值减少。[28]然而，一些看来不是问题的事情，有时候也会成为一个问题。就像通常所说的那样，偶尔会碰到一次“挫折”。可以将这种挫折视为一个有待克服的障碍，所以它会使 A 的值增加。最后，假如碰到挫折的次数足够多（A 就会变得足够大），那么就可以认为这方面的研究不会有什么成果，所以就应该转而研究其他技术。

可以将上述随机过程简单地设定为扩散过程。0 为吸收壁（absorbing barrier），可以将其理解成“发现”；$\bar{A}$ 是反射壁（reflecting barrier），它会使企业从事新的研发活动。[29]

因此，从数学的角度看，对于任何一个特定的扩散过程，我们的问题就是去描述最先通过 A 到 0 这段时期的分布。在我们开始展开分析时，将会把研究单位的数目看成是外生的。接着，我们再将研究单位的数目视为内生决定的。

首先，考虑一个可以在研发领域展开竞争的市场。假定有 N（$\geqslant 2$）个实验室（企业）在相互独立地工作。为简化分析，我们假设这些实验室都是相同的。令 $h(t, \alpha)$ 为研究计划 α 最先通过的密度函数，$H(t,\alpha) \equiv \int_0^t h(\tau,\alpha)d\tau$ 为累积函数。与前面一样，我们用 V_e 来表示一个成功的研究项目对企业的资本化价值。于是，对代表性企业而言，研究计划 α 的期望现值就是

$$V_e\int_0^{\infty} e^{-rt}h(t,\alpha)(1 - H(t,\alpha))^{N-1}dt$$

28 可以将时间间隔 A 看成是对发现的时间间隔的“主观”认识。

29 对这一扩散模型进行更为详细的数学论证，参见 Brock，Rothschild and Stiglitz（1979）。

（h 是它在 t 时刻发现新技术的概率，$(1-H)^{N-1}$ 是在 t 时刻之前它的竞争对手都没有发现该项新技术的概率）。因此，在市场均衡状态下，代表性企业会选择 α 以使得

$$\int_0^\infty e^{-rt} h_\alpha(t,\alpha)(1 - H(t,\alpha))^{N-1} dt = 0 \qquad (36)$$

此外，假设这 N 个企业由社会统一管理。令 V_s 为一个成功的研究计划的资本化社会价值。

如果存在 N 个独立的研究实验室，那么在 t 时刻发现新技术的概率为

$$Nh(1 - H)^{N-1}$$

所以，该"研究计划"的现值就是

$$W_s = V_s N \int_0^\infty e^{-rt} h(1 - H)^{N-1} dt \qquad (37)$$

因此，社会计划者就会选择 α 以使得

$$\frac{\partial W_s}{\partial \alpha} = V_s N \int_0^\infty e^{-rt} [h_\alpha(t,\alpha)(1 - H(t,\alpha))^{N-1} - h(t,\alpha)(N-1)H_\alpha(t,\alpha)(1 - H(t,\alpha))^{N-2}] dt = 0 \qquad (38)$$

由于我们想确定市场选择有风险的研究项目的水平，所以就假设，选择是从一族随机过程中做出的。对于那一族随机过程而言，发现的期望日期都相同。也就是说

$$\int_0^\infty th(t,\alpha)dt = \text{常数}$$

假设 α 的增加表示分布 $h(t,\alpha)$ 的平均保留区域，即对于所有的 $t>0$，都有

$$\int_0^t H_\alpha(\tau,\alpha)d\tau \geqslant 0 \quad t > 0 \qquad (39)$$

[30]

$$\int_0^\infty H_\alpha(\tau,\alpha)d\tau = 0 \qquad (39a)$$

此外，我们假设，在相关区域，对于所有的 α，都有

$$H(0,\alpha) = 0, H(\infty,\alpha) = 1 \qquad (40)$$

30　更多的有风险的研究项目，就会使得迅速获得结果的可能性增加，但是同时也使得完全失败（或在经过很长一段时间才能获得结果）的可能性增加。下面这个例子也许能有助于读者理解这一点。将研究过程视为一个离散的随机游走。一个研究计划就是一系列试验。每个试验的结果，或者是使研究者更接近其目标，或者是根本不能带来信息。尽管一个小试验也许只能使研究前进"一步"，而一个大试验也许就能使研究前进"两步"，但是它失败的概率也更高。因此，一方面，对于有风险的研究计划而言（假如所有试验都能成功），它成功的最短时间是其他研究计划的一半。另一方面，由于试验不成功而使研究期间变得更长的概率也更高。

现在，我们就可以证明，当且仅当 $r=0$ 时，α 的市场均衡选择才是最优的。如果 $r>0$，总是可以通过使 α 的值高于其市场均衡水平，来增进社会福利。

为了阐述这一点，我们对式（36）进行分部积分，并运用式（39a），就可以得到

$$\int_0^{\infty} e^{-rt}[(N-1)h + r(1-H)]H_{\alpha}(1-H)^{N-2}dt = 0 \tag{36a}$$

再次对式（36a）中的第二项进行分部积分，并运用式（39a），我们就可以得到

$$\int_0^{\infty} e^{-rt}H_{\alpha}(1-H)^{N-1}dt = \int_0^{\infty}[\int_0^t H_{\alpha}d\tau][(N-1)h + r(1-H)]e^{-rt}(1-H)^{N-2}dt \tag{41}$$

将式（36）、式（36a）和式（41）代入式（38），并令 $\alpha=\alpha_e$ 表示 α 的市场值，我们就得到，当 $r\geqslant 0$ 时，有

$$\left(\frac{\partial W_s}{\partial \alpha}\right)_{\alpha=\alpha_e} = rV_sN\int_0^{\infty}\left[\int_0^t H_{\alpha}d\tau\right][(N-1)h + r(1-H)](1-H)^{N-2}e^{-rt}dt \geqslant 0$$

$$\text{或} r \geqslant 0 \tag{42}$$

因此，如果 $r=0$，那么市场均衡就是最优的（至少局部是最优的）。然而，当 $r>0$ 时，α 的增加（表示企业承担的风险增加了）将会增进社会福利。[31]

尤其要注意到，尽管我们假设，每家企业只能从一些研究策略中做出一个选择（在所有研究策略中，发现的平均时间都相等），但是对于整个市场而言，发现的平均时间是会随着 α 的变动而变动的。为求简化，我们只考虑 $N=2$ 的情形，连续进行分部积分，并利用式（39），就可以得到

$$\bar{T} = 2\int_0^{\infty} th(1-H)dt \tag{43}$$

$$\begin{aligned}\frac{d\bar{T}}{d\alpha} &= 2\int_0^{\infty} t[h_{\alpha}(1-H) - hH_{\alpha}]dt \\ &= -2\int_0^{\infty}[H_{\alpha} - \int_0^t (Hh_{\alpha} + hH_{\alpha})d\tau]dt \\ &= 2\int_0^{\infty} HH_{\alpha}dt \\ &= -2\int_0^{\infty} h[\int_0^t H_{\alpha}d\tau]dt < 0\end{aligned} \tag{44}$$

---

31 由于 W_s 通常都不是一个性态良好的凹函数，所以也可能会出现，在全局最优状态下，企业承担的风险水平更低这种情况。

因此，在市场经济中，发现的平均时间要低于社会最优水平。

上述分析研究了，在竞争性市场和社会管理的市场上，企业所承担的风险的性质。在这两个市场上，外生给定的研究实验室的数目是相同的。现在，我们允许 N 是内生的。于是，假设 F 表示建立一个实验室的"固定成本"。

显然，现在就有

$$W_s = NV_s\int_0^\infty e^{-rt}h(1-H)^{N-1}dt - FN \tag{45}$$

对 N 微分，我们就能得到

$$\frac{\partial W_s}{\partial N} = \frac{W_s}{N} + NV_s\int_0^\infty e^{-rt}h(1-H)^{N-1}\ln(1-H)dt - F = 0 \tag{46}$$

由于存在自由进入，所以就可以用企业数目 N_c 来表征市场均衡。N_c 是下式的解

$$V_e\int_0^\infty e^{-rt}h(t,\alpha)(1-H(t,\alpha))^{N-1}dt = F \tag{47}$$

比较式（46）和式（47），我们发现，市场均衡状态中的企业数目既可能会高于社会最优水平，又可能会低于社会最优水平。这里有三个效应：

（1）市场通常都不能获得所有的研究收益，即 $V_e < V_s$。这会使得市场中从事研发的企业不足。

（2）$\int e^{-rt}h\ (1-H)^{N-1}\ln(1-H)\ dt<0$（由于 $H\leqslant 1$，$\ln(1-H)\ \leqslant 0$）。社会计划者会考虑到，另外增加一个研究实验室，就会使其他企业的研究收益减少；而市场就会忽视掉这一影响（这与前一部分中的结论是一致的。也就是说，在具有自由进入特征的市场均衡中，平均收益等于平均成本；而社会最优则要求边际收益等于边际成本。由于边际收益要低于平均收益，因此市场经济中就可能会有过多的企业从事研发活动）。

（3）在选择研究策略时，互不协调的行为（在此反映在 $\alpha_e < \alpha_s$ 这一事实上）会使得市场均衡状态下的研究计划之价值，低于社会最优状态下研究计划的价值。即使在存在创新收益的专属性时也是如此。这种效应可能会使市场均衡状态下的企业数目，小于社会最优状态下的企业数目。[32]

与这些研究结论相对比就可以发现，一家控制着所有研究实验室的垄断企业，总是能够做出社会最优的技术选择（$\alpha_m = \alpha_s$）。但是，因为 $V_m < V_s$，

32　然而，假如政府不能改变市场关于 α 的选择，那么政府是否想增加或减少研究单位数目（比如通过对分支机构征收数额为 F 的税收），只取决于前面两种效应。

所以由垄断企业建立的研究实验室的数目肯定会低于社会最优水平。因此，在垄断情形下，出现创新的平均时间要长于社会最优水平。

六、结论

在本文中，我们研究了研发活动中竞争的性质和后果，以及研发活动中的竞争与产品市场上的竞争的关系。过去的一些研究是为了确定：垄断是否会比竞争导致更多的研究？或者，与社会最优水平相比，竞争性环境下的研究到底是多了还是少了？这些研究在两个关键方面都受到了限制：它们没有区分当前产品市场上的竞争和研发活动中的竞争，而且它们也没有认识到市场结构本身就是一个内生变量。在本文中，我们主要探讨了以下四个问题：

（1）如果可以在研发领域展开竞争，那么当前的市场结构（目前产品市场上的竞争状况）对研发活动的影响是什么？我们在本文中指出，与当前的市场结构是竞争性时的情形相比，当前的市场由一家垄断企业主导时，研发支出会更多。原因仅仅在于，垄断企业支付更多的研发成本，那么在创新出现之后的市场上，竞争就会更少，从而利润就更多。

（2）研发活动中的竞争对研究水平的影响是什么？在本文中，我们指出，与完全垄断时的情形相比，竞争总会带来更多的研究。实际上，由于有自由进入，所以就会有平均收益等于平均成本，即使只能获得部分收益，竞争仍然会使研发支出高于社会最优水平。

（3）对于时间，研发活动中的竞争怎样影响产品市场上的竞争？也就是说，如果在研发活动中没有进入壁垒，那么研发活动中的竞争是否会导致新企业进入产品市场，从而使垄断企业的存续期变短？我们的分析表明，在一定条件下，垄断会持续下去：如果在产品市场上，目前的垄断企业也能获得研发技术，它就能（并且将会）阻止其他企业进入。例如，它可以通过进行足够快的研究，来使其他开展研发活动的进入者得不到任何收益。在这些条件下，尽管竞争威胁也许会使垄断企业开展足够快的研究（这个速度要大于没有竞争威胁时的速度），但是垄断企业的存续期并不会缩短。

因此，即使在任何一种市场情形中，实际上没有竞争者，但是潜在的竞争可能会产生重要影响（当然，这与一个广为人知的发现有关。即一个产业中的企业数目或者产出在这些企业间的分布，并不能很好地度量该产业的竞争程度。一个极端情况就是，在第三部分的模型中，研究并不是随机的，这样每次都只有一个企业开展特定的研发活动）。

在第三部分的模型中，任何企业只需要增加研发支出，就能将发现日期

提前。也就是说，更多的研究与更快的研究是等价的。在这些模型中，没有不确定性（或者说，即使有不确定性，它也是非常简单的；而且所有企业都只能采取一个研究策略）。从更快的研究中获得的收益，就是在 $t-\Delta\hat{t}$ 时刻（而不是在 t 时刻）做出创新所带来的收益的*增加量*（以现值计算）。然而，假如是一家企业做出了创新，那么私人收益就是专利权的*全部*价值。在非随机模型中，若一个研究者的研究速度比最快的速度稍微慢一点，则其研究的社会收益和私人收益都等于零：它的努力纯粹是重复性浪费，社会也不会向它提供任何报酬。如果研究结果并不完全相关，那么情形就不是这样的了。

（4）是哪些因素决定了在任何时候都会从事研发活动的企业（实验室）的数目？在本文中，我们注意到，研究结果的不确定性所起到的关键作用：假如没有不确定性，那么就只有一家企业从事研发活动；如果存在不确定性，可能就有几家企业开展研发活动（更精确地说，不同企业所采取的研究策略的*相关性*是非常关键的。在本文中，我们考察了两个极端的情形，即各自的研究结果是完全相关的这种情形，以及各自的研究结果是独立分布的这种情形。正如我们已经强调的那样，应该将相关程度本身视为一个内生变量。我们希望在其他文章中研究这一问题）。

此外，我们比较了竞争性市场均衡中的企业数目和社会最优的研究单位数目。由于从研发活动中获得的收益是不完全相关的，所以每个研究单位的边际社会收益就为正。但是，边际收益要低于平均收益。在市场均衡状态下，（*平均的*）私人收益就等于实验室的成本。然而，社会最优状态要求*边际社会收益*等于实验室的边际成本。因此，研究单位数目是多了还是少了，取决于需求弹性。需求弹性决定了私人收益与社会收益的比率，以及创新的规模。[33]另外，我们的分析表明，如果有足够多的企业，以至于能使每家企业可以忽视其对发现日期的概率分布的影响，那么每个研究单位都会在有效的水平上运行。然而，假如各企业不能忽视这种影响，每个实验室就会在*无效的*水平上运行。

市场均衡状态不仅在研究实验室的*数量*上，而且还在项目风险上不同于社会最优状态。[34]在本文中，我们注意到，如果市场利率高于零，则市场均衡会系统性地偏向于过低的风险。相比较而言，垄断企业总是会有效地利用每一个研究实验室（包括规模和风险选择），但是一般来说，它们建立的研

33 当需求弹性是一个常数时，这也是对的。如果要在一个略为不同的背景下讨论，当需求弹性不是常数时分析会如何变化，那么需求弹性为常数就是一个关键性假设。

34 还有另外一种情况，在此情形下，市场均衡状态下的研究策略会不同于社会最优的研究策略。例如，我们已经指出的不同企业所采取的研究策略之间的相关程度就是这种情况。

究实验室数量过少。

因此，有限竞争是否会比自由竞争引致更多研究，是一个非常复杂的问题。一方面，如果没有占有收益的方法（即某种程度的垄断），例如专利，那么就没有任何企业会进行研发。另一方面，研发竞争会激发研发投入，但这种研发投入的效率低于垄断状态。这样，与垄断状态相比，研发竞争可能会延迟做出发现的日期。[35]尽管垄断状态很可能导致研发投入不足，但（有专利权的）竞争市场中的研发投入却可能过多。

最后，我们的分析表明，那种认为研发竞争可以代替或最终引致产品市场竞争的观点，是值得怀疑的：在某些条件下，只要存在由专利制度（或信息扩散的滞后）形成的进入壁垒，即使再也没有任何其他正式的进入壁垒，垄断状态依然可以持续。

无论如何，我们至少希望可以使读者相信，在技术变迁比较重要的那些经济部门，以传统的竞争均衡模型（阿罗—德布鲁模型）为基本框架的市场竞争均衡分析，其用途是有限的：研发竞争一定是在产品市场不完全竞争的情况下出现的。这种类型的竞争需要一个完全不同的分析方法。

参考文献

ARROW, K.J. "Economic Welfare and the Allocation of Resources for Invention" in R. Nelson, ed., *The Rate and Direction of Inventive Activity: Economic and Social Factors*, NBER, Princeton University Press, 1962.

BARZEL, Y. "Optimal Timing of Innovations." *Review of Economic Studies*, Vol. 35, No. 3 (1968), pp. 348–55.

BROCK, W., ROTHSCHILD, M., AND STIGLITZ, J.E. "Notes on Stochastic Capital Theory." Mimeo, Oxford University, 1979.

DASGUPTA, P., GILBERT, R. AND STIGLITZ, J.E. "Invention and Innovation under Alternative Market Structures: The Case of Natural Resources." Mimeo, Oxford University, 1979.

——— AND HEAL, G. in J. Nisbet, ed., *Economic Theory and Exhaustible Resources*, 1979.

——— AND STIGLITZ, J.E. "Market Structure and Research and Development." Mimeo, London School of Economics, 1977.

——— AND ———. "Industrial Structure and the Nature of Innovative Activity." *Economic Journal* (1980).

DIXIT, A. AND STIGLITZ, J.E. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review*, Vol. 67 (1977), pp. 297–308.

GALBRAITH, J.K. *Economics and the Public Purpose*. 1973.

GILBERT, R. AND NEWBERY, D. "Preemptive Patenting and the Persistence of Monopoly." Mimeo, Cambridge University, 1979.

——— AND STIGLITZ, J.E. "Entry, Equilibrium, and Welfare." Paper presented to a conference on Theoretical Explorations in Industrial Organization, Montreal, October, 1979.

35 Gilbert and Stiglitz（1979）探讨了一种情形。在该情形中，竞争越激烈，创新出现的日期就越迟。

KAMIEN, M. AND SCHWARTZ, N. "Market Structure and Innovation: A Survey." *Journal of Economic Literature*, Vol. 13 (1975), pp. 1–37.

LOURY, G. "Market Structure and Innovation." *Quarterly Journal of Economics*, Vol. 93 (1979), pp. 395–410.

NELSON, R. AND WINTER, S. "Factor Price Changes and Factor Substitution in an Evolutionary Model." *Bell Journal of Economics*, Vol. 6, No. 2 (Autumn 1975), pp. 466–486.

SALOP, S. "Monopolistic Competition Reconstituted, or, Circular Fashions in Economic Thought." Federal Reserve Board, Special Studies Paper No. 89, Washington, 1977.

———. "A Note on Self-Enforcing Threats and Entry Deterrence." *American Economic Review*, Vol. 69 (1979).

SCHERER, F.M. *Industrial Market Structure and Economic Performance*. Chicago: Rand McNally & Co., 1970.

SCHMOOKLER, J. "Economic Sources of Inventive Activity." *Journal of Economic History*, Vol. 22 (1962).

SCHUMPETER, J. *Capitalism, Socialism, and Democracy*, 2nd ed. London: Allen & Unwin, 1947.

SCHWARTZMAN, D. *Innovation in the Pharmaceutical Industry*. Baltimore: Johns Hopkins University Press, 1977.

SPENCE, A.M. "Product Selection, Fixed Costs, and Monopolistic Competition." *Review of Economic Studies*, Vol. 43 (1976), pp. 217–236.

———. "Entry, Capacity, Investment, and Oligopolistic Pricing." *Bell Journal of Economics*, Vol. 8, No. 2 (Autumn 1977), pp. 534–544.

STIGLITZ, J.E. "Perfect and Imperfect Capital Markets." Mimeo, Yale University. Paper presented at the Winter Meeting of the Econometric Society, 1971.

———. *Information and Economic Analysis* (forthcoming).

WILSON, R. "Informational Economies of Scale." *Bell Journal of Economics*, Vol. 6, No. 1 (Spring 1975), pp. 184–195.

技术变迁、沉没成本与竞争*

在广为人知的关于资本主义优点的观点同我们用以证明其优点的模型之间，长期存在着一种分歧。我们之所以相信资本主义的经济组织形式，主要是因为它的动态特性，以及它带来的生活标准的提高，但是我们的经济理论——至少我们教给学生的经济理论——是以假设技术不变的模型为基础。

这种分歧很让人烦忧，不仅仅是因为它会导致知识上的不协调。关于重要政策问题的观点往往都基于过于简化的模型，无视这些模型有多么不合适。比如，在考虑对反托拉斯政策（包括旨在减少对贸易垂直限制的政策）的需求及其后果的时候，通常都是以标准的竞争模式为基础。再比如，人们认为，生产者只有在限制会增进效率的时候，才会实行垂直限制。然而，技术变迁相当重要的市场从来不是完全竞争的，并且在不完全竞争的市场上，垂直限制可能会改变竞争的程度。[1]

在资本主义、社会主义以及民主政治中，约瑟夫·熊彼特早已指出，任何关于资本主义的分析都必须以存在技术变迁的模型为基础，在这些模型中竞争是以新产品或者新工艺的开发存在，而不是（或者同时）以传统价格理论关注的价格竞争的形式存在。他指出，带有技术变迁的市场天然就是不完全竞争的。尽管熊彼特意识到价格竞争的相对不重要性，但他并没有充分地进行说明。

对于传统竞争理论为何不能很好地解释工业部门的情况有多种原因。在工业部门技术变迁中扮演着重要作用。其中的一种作用是：技术变迁自然地使得

---

* 非常感谢 Partha Dasgupta，Steve Salop 以及 Richard Gilbert 与我的有益探讨。同时感谢 Mark Schwartz，Andrew Postewaite，Richard McMasters，Hal Varian 以及布鲁金斯大会（Brookings Conference）的参加者们对本文做出的评论。本文的研究得到了美国国家自然科学基金会、俄林基金会（Olin Foundation）以及斯坦福大学胡佛学院（Hoover Institution，Stanford）的资助。

1 在垄断市场上会采用垂直限制，这或者是因为它们增进了效率，或者是因为它们增强了垄断者价格歧视的能力。因此，研究垄断和竞争的极端情况可能不能对更普遍的不完全竞争的情况做出解释——在不完全竞争的情况下，采用垂直限制是为了改变竞争的程度。要了解更多关于垂直限制对改变竞争程度的作用的讨论，请参见 P. Rey and Joseph E. Stiglitz，"The Role of Exclusive Territories in Producers' Competition，" unpublished paper（1987）；以及 Rey and Stiglitz，"Imperfect Competition and Vertical Restraints，" *European Economic Review*（forthcoming）。

一种要素具有递增收益，同时使得（大部分）研发的费用成为沉没成本。

正如我们长久以来认识到的那样，在存在收益递增的时候，竞争可能无法实行：市场均衡和经济效率都只允许有一家企业，或者至多是一小部分企业存在。然而，近些年来，出现了另外一种观点：即便在那些只有有限个竞争者的市场上，也可以采取竞争行为，这是因为潜在竞争造成了纪律约束。这一观点被称为竞争性学说（contestability doctrine）。[2]正如效率所要求的那样，在具有递增收益的行业中，只会存在一个企业。该企业在价格等于平均成本的水平上实现产出的最大化——也就是说，它的利润为零（图1）。如果垄断者产出水平较低，同时价格高于平均成本，就会有新的进入者在这一市场参与竞争，新的进入者以较低价格进入，将原来企业的顾客全部转变为自己的顾客。

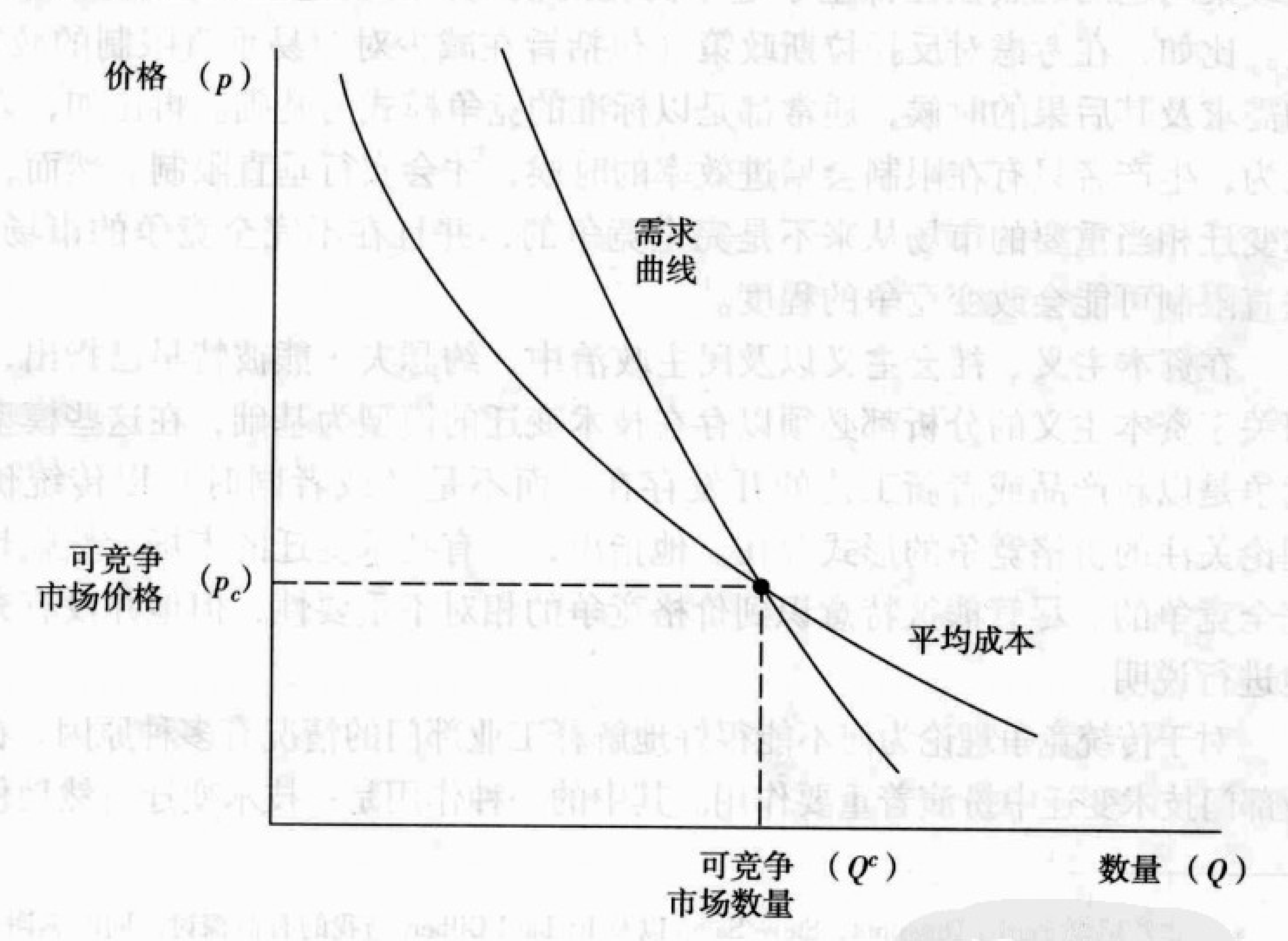


图1 竞争性市场均衡

2 关于这一观点，很明显，在“芝加哥学派”（Chicago School）中早有先行理论。比如，可参见 Harold Demsetz，“Why Regulate Utilities,” *Journal of Law and Economics*, Vol. 11（April 1968），pp. 55–66。近期最热情的拥护者包括 Sanford J. Grossman，“Nash Equilibrium and the Industrial Organization of Markets with Large Fixed Costs,” *Econometrica*, Vol. 49（September 1981），pp. 1149–1172；William J. Baumol, *Contestable Markets and the Theory of Industry Structure*（Harcourt Brace Jovanovich, 1982）。

如果竞争性学说是正确的，它将极大地改变我们对反托拉斯政策的态度。这将意味着以下这类重要的情况不需要或者不会存在：由于存在着收益递增，市场上会产生自然垄断（natural monopoly）（或者寡头垄断），而以前人们总是认为，收益递增会导致市场失败（帕累托无效）。如果这种观点是正确的，那么就不需要政府干预，即便市场上只有一个或者很少企业。事实上，由于潜在的规模效应无法得到充分的挖掘，典型的政府干预形式（拆分企业或者限制企业合并）可能是相当有害的。

传统的经济分析认为，在收益递增的条件下，存在着下列交替关系：如果市场上的企业很多，垄断的力量比较小，就会使生产效率受损；如果市场上企业比较少，垄断的力量比较强，就会使生产效率增加。但是竞争性学说认为，并不存在这样的交替。

对于那些有着领先技术的国家而言，这些问题是尤其重要的。当技术变迁是研发或者干中学的结果时，就可能存在收益递增以及市场力量的高度集中。如果潜在竞争既能保证行业运行的效率，又能保证效率提高的好处能够转移给消费者，那么政府对那些通过技术优势获得主导地位的企业的干预——拆分或限制企业大小——就是非常不明智的了。

本文的目的就是要强调"潜在竞争足以保证经济效率"这一假设的有效性和一般性。本文既给出了用以证明我们对经济政策的设计理念的模型，同时也指出了这些经济政策应该是怎样的。后一点是相当困难的，我在结论部分将列出一些观察报告。

更精确地说，我认为反对政府开展竞争政策（如反托拉斯规定、对公平交易的管制等）的传统观点，对现代经济的大多数部门而言，只有有限的有效性和相关性。这些观点认为：

——竞争确保了效率，技术变迁带来的收益以低价格的形式转移给了消费者；

——利润吸引新的进入者，确保了市场的竞争性。

竞争性学说强调，要估计竞争的力量，不应当只看市场上现有企业的数量，而应该看潜在进入者。在大多数情况下，都有很多的潜在进入者。但是正如市场上企业之间的伯川德竞争（Bertrand competition，即残酷的竞争）只需要两个企业来获得竞争结果那样，竞争性学说认为，只要有一个潜在进入者，或者最多只要很少的潜在进入者，就可以同时确保经济效率和零利润。

与之对照的是，我认为这些结果对零沉没成本这一假设是相当敏感的。

在存在沉没成本的情况下，即便只是很小的沉没成本，我认为：

——利润的存在可能不会吸引新的进入者。潜在进入者会对他们进入之后的情况进行判断，尽管已有的垄断者可以获得正利润，但试图与之在市场上竞争的进入者很有可能获得负的利润。因此，潜在竞争并不是充分条件。

——进入者并不能保证竞争。新的进入者（们）可能和现有企业串谋而不是竞争（并且这种串谋可能是暗中的而不是公开的，这由每个企业对各自利益的理解所推动）。或者，进入者可能仅仅导致某个企业的退出——而不是竞争。

——现有企业会采取措施阻止进入；进入的企业会采取措施串谋。其中一项措施是技术选择——包括决定是否采用需要很大沉没成本的技术。某些阻止进入和串谋的措施会降低社会福利。因此，存在竞争或者潜在竞争这一事实，并不能完全证明企业采取的某些行为会增进效率。

——即便竞争可以成功地将利润降为零，也不能保证效率。我们在此所关心的竞争并不能保证价格等于边际成本：对第一个制造出新产品的企业，可能会有“租”可寻。但是，要成为第一个企业，竞争可能是非常激烈的，以至于这些“租”大部分都消失了。为这些“租”而竞争，可能会赋予资本主义本质上的动态性，但并不一定导致标准的帕累托意义上的效率。

因为技术变迁内在地包含了收益递增和沉没成本，竞争性学说对于技术变迁相当重要的行业是非常不适用的。潜在竞争既不能保证经济效率，也不能保证零利润。

本文分为五个部分。我将证明在静态模型中，如果存在沉没成本（即便很小），潜在竞争并不足以保证经济效率，并且事实上沉没成本是很普遍的。接下来的部分将考虑一些重要的反对本论断的观点。我将证明它们不是有效的，或者说它们增强了人们对潜在竞争的有限效率的关注。然后，我会将这一分析扩展到创新的沉没成本问题中。本文的结尾是一些简要结论，主要是解释论证结果对经济政策的含义。

一、沉没成本和潜在竞争

在过去十年中，关于产业组织研究的主要观点是：对于进入和竞争本质至关重要的是新企业进入时企业的反应，或者潜在进入者对上述反应的看法

（这些通常都不为人知）。[3]现有企业的当前价格可能很高，但是如果潜在进入者相信进入会导致激烈的竞争和较低的价格，它就不一定选择进入。因此，在事前竞争（ex post competition）、进入前竞争（competition before entry）、事后竞争（ex post competition），以及进入后竞争的性质（the nature of competition after entry）之间存在密切的联系。事实上，过分激烈的事后竞争可能会降低事前竞争的有效性。那些决定事后竞争的性质或者说潜在进入者对事后竞争性质信念的变量，被称为状态变量（state variable）。

为验证这些问题，我假设最初在行业中只有一个企业。然后，我分两个阶段来分析：进入后的均衡（如果进入发生的话）和进入之前采取的行动，后者影响进入的可能性以及进入之后的均衡。对进入可能有三种反应：企业间进行竞争；企业可以进行串谋，分享垄断利润；或者现有企业退出。反过来，竞争也可以采取多种形式——对于双寡头竞争没有意见一致的理论。我集中分析伯川德（价格）竞争，这是因为我的目的在于证明：即便在看似最合意的条件（可以非常有效地降低价格）下，竞争也不能保证零利润或者经济效率。在本文稍后部分，我将证明在事后竞争不那么有效的时候，这些结果将产生什么变动。

因此，我们必须确知在什么情况下会产生竞争、串谋以及退出。这也正是现有企业所面临的问题。它希望首先采取行动，使市场对任何潜在进入者而言，都是没有吸引力的。它努力阻止进入。它试图说服潜在对手，一旦它们进入，就没有串谋以及利润分享之类的迁就行为，只有战争——非常激烈的竞争，从而使进入者对自己的决策后悔。图2显示了这一适时决策结构。

在本部分开始的时候，我要指出，近期关于产业组织研究的主要观点是：进入之后的竞争本质对进入至关重要，而进入之后的竞争本质由状态变量决定，比如，有约束力的承诺，以及不能挽回的投资。沉没成本（一旦

3　早期的对此有贡献的研究者包括 Steven C. Salop，"Strategic Entry Deterrence，" *American Economic Review*，vol. 69（May 1979，*Paper and Proceedings*，1978），pp. 335 - 338；Avinash Dixit，"The Role of Investment in Entry Deterrence，" *Economic Journal*，vol. 90（March 1980），pp. 95 - 106；A. Michael Spence，"Entry，Capacity，Investment，and Oligopolistic Pricing，" *Bell Journal of Economics*，vol. 8（Autumn 1977），pp. 534 - 544；Richard Gilbert and Joseph E. Stiglitz，"Entry，Equilibrium，and Welfare，" 在 NBER - NSF 会议上报告的关于产业组织的文章，多伦多，1978 年 10 月；以及 Stiglitz，"Potential Competition May Reduce Welfare，" *American Economic Review*，vol. 71（May 1981，*Paper and Proceedings*，1980），pp. 184 - 189。

图2　基本模型的适时决策结构

支出就不能偿付的成本）支出是状态变量，并且，正如我们将看到的那样，它对竞争有着重要的影响。

投资花销不必全部是沉没的。因为一家公司购买的飞机可以很容易地转卖给其他公司，对飞机的投资就不是沉没成本。同样，一家航空公司拥有的飞机的数量不能被看作状态变量。尽管许多投资都不是沉没的，但是，几乎在所有的投资里都有一种沉没成本的因素。航空公司必须通过做广告来获得消费者，它必须解决复杂的航线问题。因此，在有沉没成本，尤其是在有很小的沉没成本的时候，确认由此产生的竞争互动的本质是很重要的。

接下来的两个部分将要证明，在事后竞争非常激烈的时候，即便是很小的沉没成本也可以成为有效的进入壁垒。在两种方式下沉没成本很小：一种是相对于生产的边际成本而言沉没成本很小，另一种是“沉没”得不是很厉害——短期的无法挽回的投资似乎“沉没”得不太厉害。无论沉没成本是哪种意义上的小，潜在竞争都可能不是有效的。

1. 很小的沉没成本也能造成完全的进入壁垒

假设有两个企业，记作企业1和2。如果 a_1 和 a_2 分别是企业1和2采取的行动，则企业的利润分别记作 $\pi_1(a_1, a_2)$ 和 $\pi_2(a_1, a_2)$。企业间进行非合作博弈。假设 (a_1^*, a_2^*) 是这一博弈的唯一均衡点，并且 $\pi_1(a_1^*, a_2^*) = \pi_2(a_1^*, a_2^*) = 0$，也就是说均衡利润为零。

接下来，假设企业2可以采取被动的行动，不进入市场，这一被动行动

被记作 $\hat{a}_2$。企业 1 对 $\hat{a}_2$ 的最好反应是 $\hat{a}_1$。假设 π_1（$\hat{a}_1$，$\hat{a}_2$）>0，$\pi_2(\hat{a}_1, \hat{a}_2)$ =0。最后，假设对企业 2 而言，没有可行的行动，可以使 $\pi_2(\hat{a}_1, \hat{a}_2)$ >0。也就是说，如果第一个企业采取行动 $\hat{a}_1$，则 $\hat{a}_2$ 是企业 2 可以采取的最好反应。

现在，我们把这个博弈嵌入到一个更大的博弈中。我们将企业 1 看作现有企业，企业 2 是潜在进入者。如果企业 2 进入，则接下来的博弈就如上文所述。但是它的进入成本为 ϵ（$\epsilon>0$），同时假设这一成本是沉没的。最后假设如果企业 2 不进入，它就只能采取被动行动 $\hat{a}_2$。现在我们可以很容易地确认：

命题 1：如果 $\epsilon>0$，则存在一个唯一的子博弈完美均衡：企业 2 不进入，企业 1 选择 $\hat{a}_1$，并由此获得利润 π_1（$\hat{a}_1$，$\hat{a}_2$）>0。如果 $\epsilon=0$，则存在两个均衡：其中一个均衡是采取行动（a_1^*，a_2^*），两个企业都挣得零利润；另一个均衡是企业 2 不进入，企业 1 选择 $\hat{a}_1$。[4]

其证明是相当直接的。根据我们的假设，如果企业 2 进入了，则进入者在进入后的那一期利润为零。但是为了进入，企业 2 必须付出成本 ϵ；因此，它通过进入挣得的总利润是 $-\epsilon$。如果企业 2 选择不进入，则其总利润是零。对企业 2 而言，不进入更好。

命题 1 是令人烦扰的，因为它表明，即便一个可以忽略的进入成本也可以使现有企业免受侵扰。只要 ϵ 是正的，是否有潜在进入者在一旁觊觎就是无关紧要的。同时我们还注意到，这一命题并没有否定，现有企业自己在进入时也有沉没成本 ϵ。只要企业 1 首先行动，假如 ϵ 不是很大，它就会选择进入。并且，正如命题 1 判定的那样，企业 2 会发现进入是没有利润的。

这一讨论故意没有用经济术语来解释企业的行为。这些行为可能包含与顾客签订合约来提供商品。假设进入竞争的成本是沉没成本，命题 1 就是不受影响的。接下来的讨论，对命题 1 提供了一个具体的经济例证。

假定有一个同质产品的市场。市场需求曲线向下倾斜，单位生产成本是常数 $C\geqslant0$。企业 1 和企业 2 之间进行伯川德竞争。每个企业视另一企业的价格是给定的。伯川德竞争意味着企业将采取残酷的行为，价格将降为同边际成本 C 相等的水平。我特别关注伯川德竞争，不仅是因为它很好地描述了特定市场上的竞争——近期关于竞争性市场的模型假设这样的激烈竞争在

4　需要注意，对第 2 个企业而言，这两个均衡是无差异的，但是对第 1 个企业而言，显然不是无差异的。

实证上是很重要的——而且因为我先前所说的，伯川德竞争对竞争确保了效率和零利润这一学说提供了最好的环境（这也正是人们已经想到的）。

现在，假设进入行业需要沉没成本 $\epsilon>0$，企业1和企业2序贯进入。令 π 表示不计进入成本的行业垄断利润。这样，容易确认：

命题2：如果 $0<\epsilon<\pi$，则存在一个唯一的子博弈完美均衡：只有企业1进入并且赚得纯利润 $\pi-\epsilon$。[5]

我们再次注意到，即便 ϵ 很小，企业2也不会对企业1造成任何威胁；其结果是好像没有潜在竞争。这个结果是令人烦扰的，因为我所假定的技术与规模收益不变的标准假设差别不大。由很小的沉没成本产生了一个具有递增回报的微小因素。竞争性学说认为，即便收益递增的程度很小，市场上也只能容纳一个企业，价格稍微高于边际成本；而我的分析表明，价格应当等于垄断价格。

这些分析的政策意义如何呢？任何使新的进入者进入行业的政策都会带来社会福利的改善。有两种政策证明了这一点。一种是补贴进入。传统的观点反对这种补贴政策，它们认为如果进入对社会是有益的，那么它对个人来说必然也是有益的，因此补贴是没有必要的。然而，这一观点是错误的。进入者迫使市场价格降到边际成本水平，但是消费者获得了社会收益。

第二种政策是最低价格管制（零售价格的保持）。令 $D(q)$ 表示在价格 q 下的市场需求。现在，令

$$\tilde{q}=C+[2\epsilon/D(\tilde{q})] \tag{1}$$

（见图3）。由此很容易确定：

命题3：如果政府设定一个稍微高于 $\tilde{q}$ 的最低价格，则存在一个唯一的子博弈完美均衡：两个企业都进入市场，并且采取允许的最低价格。

这一政策背后的思想非常简单。在命题2中，即便很小的进入成本也能为现有企业提供完全的保护，因为模型设定：如果两个企业都进入，则会有

---

5 为了标记上的简便，我们假设利率为零。如果企业在两阶段博弈中同时行动，其中进入先于价格竞争，则存在三个子博弈完美均衡：企业1进入，企业2不进入（如2所述）；企业2进入，企业1不进入；还有一个对称的混合策略均衡（每一个企业都以概率 p 进入），其中每一个企业的预期利润都为零。文中忽略这种对称的结果，因为我希望探讨潜在进入者对现有企业行为的影响，以及对最后的行业结构的影响。只有非对称均衡是生产性有效的。有 P^2 的概率存在双倍的沉没成本，有 $(1-P)^2$ 的可能没有企业进入。当然，在不对称均衡中，由于价格超过边际成本，就这点而言是没有效率的。即便用简单的准则，比如消费者剩余和生产者剩余之和，总体的福利比较都不是很清晰的：它们取决于沉没成本的大小，以及消费者剩余的大小。

残酷的竞争。在命题3中，政府限制了残酷的竞争，从而允许两个企业都进入，并且可以赚得一定利润，因此进入是有利可图的。还需注意到，在式（1）中，如果 ϵ 很小，q 只比 C 高一点。因此，如果进入成本很小，则命题3给出的最低价格管制，维持了一个近乎有效率的结果。

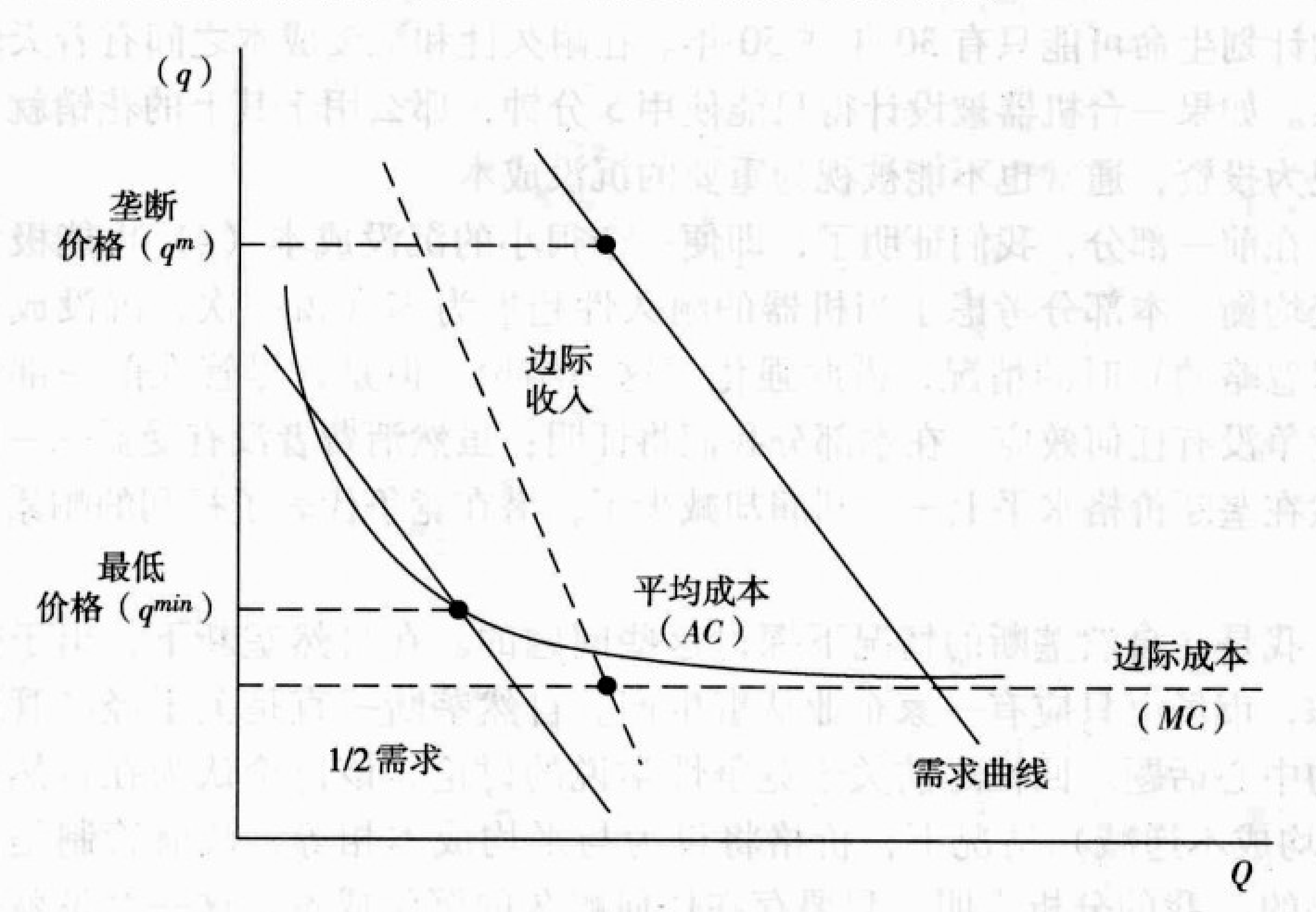


图3　零售价格的保持

我并不确信在很多种情况下，这个模型都可以对零售价格的保持提供合理解释，我将在后文解释部分原因。由于很难确定在什么情况下这些政策可以增加福利，我也不能确信这两种政策中哪一个是大体合意的。[6]

另外，这一部分的分析也是一种警告。如果在分析竞争时，关于技术的标准假设的微小改变，能够带来关于市场均衡和对政府干预的需求的巨大变化，我们对由这些分析（尤其是近来基于竞争性学说的一些变量）所产生的政策是否还那么有信心？

6　如果人们同意这是对进入补贴的合法证明，市场的潜在进入者就会认为，这样的情况是适用于他们的。类似地，如果人们同意这些政策是对零售价格保持的合法证明，行业中的企业就会认为，它们所面临的正是这样一种情况：如果没有零售价格的保持（RPM），就会存在破坏性竞争（destructive competition）。在这些政策确实合适的情况下，即便有政府制定的这些政策，也无租（利润）可寻，因此也就没有寻租行为。

在这两种政策中，补贴政策有一定优势，它只需要政府的一次性行为，不需要长期的管制。零售价格的保持政策的有效性依赖于：政府知道需求函数和技术（沉没成本的价值）。

2. 沉没成本、耐久性和自然垄断

沉没成本的本质特征是，它们代表了不能偿付的花销。它们通常表示那些用于购买不可运输的、不可转换的工厂和设备的花销，比如核电厂的遗址修复工作。尽管在最好的情况下，工厂和设备的使用期限是无穷的，但是工厂的计划生命可能只有30年或50年。在耐久性和沉没成本之间有着天然的联系。如果一台机器被设计得只能使用5分钟，那么用于其上的花销就不能被视为投资，通常也不能被视为重要的沉没成本。

在前一部分，我们证明了，即便一个很小的沉没成本（ϵ）也能极大地改变均衡。本部分考虑了当机器的耐久性趋近为零（再一次，沉没成本是可以忽略的）时的情况，借此强化了这一结论。但是，尽管在前一部分潜在竞争没有任何效应，在本部分我们将证明：虽然消费者没有受益——价格仍然在垄断价格水平上——利润却减少了，潜在竞争代表了福利的帕累托恶化。

我是在自然垄断的情况下探讨这些问题的。在自然垄断下，出于效率考虑，市场上只应有一家企业从事生产。自然垄断一直是关于经济管制讨论的中心话题。回忆之前关于竞争性学说的讨论，该讨论认为在自然垄断（平均成本递减）情况下，价格将设为与平均成本相等，政府管制是没有必要的。[7] 我的分析表明，只要存在任何耐久的沉没成本，这一结论就是无效的。

考虑这一情况，具有同质产品的市场面临需求函数 $q(Q)$，其中 Q 是在任何日期的产出流，q 是相应的市场出清价格。为了进行生产，需要建造成本为 F 的设备。这种设备的精确使用寿命是 T 年，并且它们在 T 年一次性折旧。一旦安装这一设备，就可以以边际成本 C（$\geqslant 0$）生产产品。因此没有生产能力限制。令 r（>0）表示利率。为了简便，令时间是连续的。扣除了可变成本的企业收入 R 为：

7　进一步讨论表明：如果自然垄断者制造很多种商品，它将采用拉姆齐价格（Ramsey price），正如政府想要采取的价格一样。尽管事实上，垄断者不会简单地令边际收入等于边际成本（如更简单的垄断理论中那样），并且它索取的价格和拉姆齐价格有表面的相似性，但总体而言，这一价格和政府代理人运营企业并最大化社会福利所索取的价格是不同的。要对我们探讨的问题给出例证，请参见 David E. Sappington and Joseph E. Stiglitz, "Information and Regulation," in Elizabeth E. Bailey, ed., *Public Regulation: New Perspectives on Institutions and Policies* (London: MIT Press, 1987), pp. 3 – 43。

$$R(Q) \equiv q(Q)Q - CQ^{8} \tag{2}$$

令 Q^m 表示最大化 $R(Q)$ 的值；Q^m 是在行业没有潜在进入者的情况下，垄断者会生产的产出量。技术和需求函数满足有且仅有一个企业可以随时间流逝赚到足够的利润，来弥补进入的固定成本。[9]

如果假设 F 表示并不沉没的固定成本，则竞争性市场理论断定，如果正在讨论的市场是可竞争的，则潜在进入的威胁会迫使价格降低到利润为零的水平——也就是说，企业收入的折现价值扣除可变成本正好等于固定成本，每 T 年会建造一个新设备。竞争性的产出 Q^c，由零利润条件的解给出：

$$R(Q^c) = rF/(1 - e^{-rT}) \tag{3}$$

很清楚，$Q^c > Q^m$，即竞争性的产出高于垄断产出。

现在，假设 F 表示沉没成本。假设一旦建造了设备，就存在伯川德竞争，并且产品的均衡价格等于产品的边际成本 C。如果存在潜在进入者，现有企业不会通过将产出提高到 Q^c 来阻止进入，而是通过以小于 T 的周期建造新的设备来阻止进入。

现有企业在 $T=0$ 时有沉没成本 F。其竞争对手如果立刻进入，则无法获得任何收益。因为在整个机器的使用寿命期间，伯川德竞争都确保了价格恰好等于 C，进入者无法弥补沉没成本。然而，在 T 之前存在着某个日期，如果现有企业没有建造第二部生产设备，则竞争者进入并且付出沉没成本 F 是有收益的。这是因为，尽管在两家企业同时生产的时候，进入者将赚取零利润，但是在现有企业的设备使用期结束之后，进入者就成为垄断者了。当然，如果竞争者在这一日期进入是有利的，那么对现有企业而言，在这一日期之前重新建造设备，付出沉没成本 F 就是有利的。通过预先占领对手的位置，现有企业可以阻止进入，并将价格保持在垄断价格水平。因此，为了阻止进入，现有企业每 τ 年就会建造新的生产设备，其中 $0<\tau<T$；因为垄断者不受到进入的威胁，因此它在垄断水平上生产 Q^m。

为了计算出阻止进入年份 τ，假设竞争者在 τ 年（$<T$）付出沉没成本 F 进入行业，并且宣布它将每过 τ 年就付一次沉没成本 F。如果这一计划实现了的话，现有企业就会在 T 时期离开。在 $[\tau, T]$ 区间，双寡头都不能

8 我假设 $R(Q)$ 是 Q 的凹函数。对于更一般的情况，分析会更复杂一些，但是结论不会改变。

9 这被称为强自然垄断。其条件是 $2F/(1-e^{-rT}) > R(Q_m)/r > F/(1-e^{-rT})$。如果这个式子成立，在这个行业中就只有一个企业可以生存。

获得运营利润，从 T 时期开始，进入者在每一刻都挣得 $R(Q^m)$。从获得优势的 $t=0$ 点，进入者获得的利润的现值就是

$$[R(Q^m)e^{-rT}]/r-[Fe^{-r\tau}/(1-e^{-r\tau})]^{10} \tag{4}$$

在某个唯一的 τ 值上，该表达式值为零。我们将这个唯一的 τ 值记作 τ^*。因此 τ^* 就是进入者（后来成为市场上唯一的生产者），恰好能保持收支平衡的更新设备周期。τ^* 是下式的解：

$$[R(Q^m)e^{-rT}/r]=[Fe^{-r\tau}/(1-e^{-r\tau})] \tag{5}$$

这一模型的静态子博弈完美均衡是：现有企业每 τ^* 年重新建造生产设备。一旦现有企业延迟重建设备，它就将市场拱手让给了对手，因为如果现有企业没有在 τ^* 之前建第二套生产设备，对手进入就能获得收益。一旦对手进入市场并付出沉没成本 F，现有企业就会发现再建设备不再有利可图，它的处境变得和其他的潜在进入者的处境相仿。现有企业了解到这一点，就会被迫每 τ^* 年重新建造设备。如果它这么做，就没有对手会进入。因为 $\tau^*<T$，要阻止进入就产生了一定的成本：在现有企业重新建造设备的时间 τ 和设备自然淘汰的时间 T 之间，有一段空当，现有企业在这段时间内必须让旧设备闲置。

在没有竞争的时候，现有企业利润的贴现值是：

$$[R(Q^m)/r]-[F/(1-e^{-rT})] \tag{6}$$

我们已经认识到，竞争的威胁使现有企业的利润降低到

$$[R(Q^m)/r]-[F/(1-e^{-r\tau^*})] \tag{7}$$

潜在竞争丝毫没有影响到消费者的利益，它仅仅造成现有企业利润的减少。因此，潜在竞争导致了帕累托较差结果。

尽管利润降低了，但它并没有降为零。因为 $\tau^*<T$，比较式（5）和式（7），可以发现除一种受限情况外，现有企业阻止进入的策略产生了正的利润。如果利率很低，式（5）简化为 $R(Q^m)\ \tau^*\approx F$，并且在设备使用期间（在它被新设备取代之前）获得的利润，必然正好等于固定成本。由此，被迫每 τ^* 年更换设备的现有企业，必然几乎无法获得任何利润。但即便是这样，这也同竞争性市场的结果完全不同。每个投资周期获得的利润都消散了，不是因为将产出扩大到了 Q^c，而是因为过于频繁地建造生产设备。

10　第一项是不变运营利润流 $R(Q^m)$ 的贴现值。该利润流始于现有企业离开该行业的 T 时期。第二项是无穷几何级数的加总，$F\ [e^{-r\tau}+e^{-2r\tau}+e^{-3r\tau}+\ldots\ e^{-nr\tau}]$，代表每 τ 期用于工厂的花费的贴现值。

因此：

命题4：进入的威胁迫使现有企业过于频繁地重建生产设备，超出了必要的频率。现有企业不是通过扩大产出，而是通过在一定时期内保持闲置生产力来阻止进入。潜在竞争对于社会福利是有害的。现有企业获得的利润仍然是正的，除了在零利润这一受限情况下。

这一模型有明显的政策意义。竞争，无论是真实的还是潜在的，都不足以保证自然垄断的效率。政府管制，包括特许经营权等对进入的限制，都是必需的。[11]

现在，我们要讨论当沉没成本沉没得不太厉害的时候，会怎样？也就是说，随着设备的耐久性趋向于零的时候，会怎样？为获得一个有意义的比较，我们假设对固定（沉没）成本花销的贴现值仍然同以前一样（如果机器在破损之后被替换）。令 $F(T)$ 表示一种技术带来的固定成本，该技术生产的机器能使用 T 年，假设：

$$F/(1-e^{-rT}) = k \text{ 对所有 } T \text{ 成立。}$$

代入式（5），我们可以得到 $\tau(T)$，表示生产上一台机器和生产下一台机器之间的时间跨度，它是技术 T 的函数。从而，$\tau(T)$ 是下式的解：

$$[R(Q^m)e^{-rT}]/(1-e^{-rT}) = (kre^{-r\tau})/(1-e^{-r\tau})$$

很清楚，随着 T 趋向于零，τ 也趋向于零。对于很小的 t 而言，

$$\tau/T \approx kr/R(Q^m)$$

这一比率并不趋向于零。价格和垄断价格相同，利润趋向于零，这是因为每一周期的有效利率都非常接近零；但是由于对生产设备重建的过分投资，所有的潜在垄断利润都消散了。尽管在工厂上的花费在一个月间沉没，但这正是现有企业所需要用来维护垄断地位的投入。潜在竞争只是迫使产生了建造第二个工厂（比如在第一个工厂建成两周之后）的沉没成本（潜在进入者也需要支付的成本）。以这种方式使成本沉没得不那么厉害，只是增加了潜在竞争的无效率。消费者的福利并没有得到提高。

3. 退出和内生竞争

事后竞争的强度，对于确定事前竞争的有效性非常重要：事后竞争越激

11　我们需要提出一个警告，这和前一模型的政策意义讨论中所注意到的一样：政府可能很难确知这一模型适用的情况。也就是说，政府很难知道，在什么样的情况下，进入限制——带有价格管制——是合意的。有一些行业可能试图受保护以免于竞争，即便这些限制不能提高福利——只要他们相信价格管制不能有效到使利润下降为零的程度。

烈，事前竞争就越无效。

当企业进入一个市场的时候，进入者和现有企业可能会串谋，而不是竞争。或者，很可能现有企业会退出。无论在哪种情况下，进入都不会使消费者享受低价。如果串谋（无论是暗中的还是公开的）被预期到的，并且确实发生了，进入就是受到鼓励的。由于过度进入，进入产生的有限竞争会使资源消散掉。消费者可能会获得稍微的改善。

再一次地，沉没成本对于确定均衡的性质非常重要。请记住，沉没成本可以被看作无法通过退出而偿付的成本。事实上，可能还有附加成本，比如说在退出的时候产生的处理垃圾的成本。

现有企业和进入者的技术选择会影响到是否会有退出、串谋或者竞争。有些时候，各种力量可能是彼此冲突的。在退出成本很高（很高的沉没成本）时，更有可能出现串谋，而这会吸引进入者。如果现有企业降低沉没成本，当对手进入市场的时候，现有企业的退出动机就增加了，这同样也会吸引新的进入者。现有企业可能不能阻止进入，但是进入并不可能有效地保证竞争。退出后存在利润这一事实并不比进入前利润必然导致进入这一事实更让退出企业后悔不迭：相关的问题并不是当前的利润水平是多少，而是利润应当是怎样的？退出之后的利润可能是正的，但是如果现有企业不退出的话，利润就会是零，正是这个促成了退出。

迫使现有企业退出。像以往一样，我们假设进入者和现有企业有同样的生产边际成本，并且在进入之后，产生伯川德均衡。令 F_e 表示进入者在退出时的所得（它所偿付的原始投资）。[12] 如果进入者进入了市场，在 $F_e > 0$ 时，现有企业退出是有利的。只要现有企业在退出的时候能够得到一定收益，就存在一个完美均衡：进入者进入市场，现有企业退出。

当然，还有另外一个完美均衡，即进入者退出；以及一个混合策略均衡，即每一个企业都以一个给定的概率退出。按照我对这一问题的设定，个人无法在不同的完美均衡之间做出选择。然而，我们假设在沉没成本上的花销越大，生产成本越低。从而，无论现有企业选择什么样的技术，进入者都会选择这样一种技术：它使得生产成本稍微低一些（也就意味着更高的沉没成本）。因此，在伯川德均衡中，进入者在可变成本上会获得微小的正回报。如果假设要维持一个生产线（无论它是否生产）至少总有微小的成本，并且记这些成本的贴现值为 C/r，只要 $C/r + F_e > 0$，唯一的完美均衡将是：

12 如果企业在退出的时候有成本，则 F_e 是负的。

最初的现有企业退出。[13]

垄断的持续和现有企业优势地位的持续。初始现有企业会离开市场，这需要和帕萨·达斯古普塔（Partha Dasgupta）以及约瑟夫·斯蒂格利茨关于垄断持续性的结论进行对比，后者表明现在的垄断者可以继续保持垄断地位。[14]这里，我已经描绘了一些非常不同的情况，其中进入的企业可以选择技术，而且现有企业无法维持其垄断地位。后来者往往拥有优势。这个结果似乎更加符合熊彼特的观点——一连串垄断企业在各自的市场上占据优势。

尽管与达斯古普塔和斯蒂格利茨的分析相比，竞争似乎是更可行的，但竞争并不能保证效率（或者零利润）。价格保持在垄断价格水平，并且进入可能导致用于沉没成本的不必要的（从社会角度来看）花销。[15]这些花销都是以进入者从现有企业手中“窃取”利润的方式实现的。

更多的潜在进入者可能导致比较无效的竞争。先前我曾经指出，事后竞争越有效，潜在竞争就越无效。存在这样一些情况：进入越容易，潜在竞争的有效性就会越小。

对于这个看似矛盾的悖论，有一种简单的解释。潜在进入者只有在相信他们可以挣得利润的时候，才会进入。但是，如果他们相信利润会迅速地被接下来的进入者窃取，他们也可能不会进入。不只是现有企业阻止了进入，潜在进入者本身也阻止了进入。进入行为受到寻求垄断租值这一动机的促使。

刚才展示的模型说明了这一观点。假设存在很多潜在进入者，但是为了简化，每一期只有一个竞争者可以进入。每一个潜在进入者都知道会有后来的进入者窃取它们占有的市场。如果进入的固定成本（除去在退出时可以偿付的成本），大于进入者占据市场那一期的运营利润的最大价值，进入就是没有收益的。

现有企业了解到这一点，就可以采取一个合适的策略来阻止进入。令$N(C)$表示具有可变成本C的技术，带来的不可偿付的沉没成本的价值。[16]令

---

13 附录A给出了另一种更加简单的构建退出—进入博弈的方法。其中，进入威胁对价格没有影响，这或者是因为不会有进入，或者是因为在进入之后，现有企业立刻就会退出。

14 Partha Dasgupta and Joseph E. Stiglitz, "Uncertainty, Industrial Structure, and the Speed of R&D," *Bell Journal of Economics*, vol. 11 (Spring 1980), pp. 1-28.

15 因为进入者选择了边际成本稍低的技术，然而，在进入之后价格比原先稍微低一些。如果现有企业在退出的时候无法偿付所有的初始投资，通常就会导致在沉没成本上的不必要的花销。

16 也就是说，如果企业在进入的时候产生成本ϵ，并且在一期后退出可以偿付F_e，那么，$N=\epsilon-[F_e/(1+r)]$，其中r表示利率。

$R(Q,C)$ 表示边际成本为 C，产出为 Q 时的运营利润，并令 $Q^m(C)$ 表示对应于边际成本 C 的垄断产出。现有企业可以通过选择一个可变成本足够低的技术，来阻止进入。这一阻止进入的技术是下列式子的解。

$$R[Q^m(C^d), C^d] = N(C^d) \tag{8}$$[17]

现在，潜在进入轻易地扭曲了现有企业的生产决策。现有企业选择可变成本比较小的技术，从而导致比较低的价格——但是，这个价格并不是竞争价格，利润也不为零。事实上，其所选择的技术一般来说都是无效率的。[18]

沉没成本很小的受限情况。现在，我们考虑沉没成本趋近于零的受限情况。同时，我考虑潜在进入者增加的时候，会出现什么情况。为了做到这一点，我稍微地调整了模型，将离散时间模型调整为直接的连续时间模型。令 ϕ 表示在每单位时间里潜在进入者进入的概率。C 取什么值时，进入的预期收入正好为零？现在假设没有对进入的阻碍。预期的净利润贴现值是 $\{R[Q(C),C]/r\}-\epsilon$。现有企业通过降低 C 来降低上述净利润贴现值，并且降低自身的利润。如果 C 下降的程度够大，利润将为零，进入将会被阻止。我们将可以满足这一条件的技术标记为 C^d。因此，第一家企业采取阻止进入的策略所获得的预期利润（不扣除进入成本）正好是 ϵ。

或者，现有企业可以允许进入。其利润（不扣除进入成本）的贴现值是：

$$\{R[Q(C^m), C^m] + \phi F_e\} / (r+\phi)$$[19]

17　本方程给出的阻止进入的技术假设进入只能维持一期。如果进入永久维持下去，则相应的方程是 $R[Q^m(C^b), C^b]/r = \epsilon(C^b)$，其中 $\epsilon(C)$ 表示可变成本为 C 的技术的进入成本。如果在退出的时候可以偿付一些成本，并且利率小于1，则 $N(C) < \epsilon(C) > r\epsilon(C)$。因此，在更普遍的情况下，阻止进入要求选择 $C = \min[C^d, C^b]$。

18　有效率的技术应当是在均衡产出的水平上，使得成本的贴现值最小化的技术。也就是说，如果 $\epsilon(C)$ 是可变成本为 C 的技术的进入成本，通过选择 C 来 $\min\epsilon(C) + (CQ/r)$。在相关的情况下，满足这一条件的可变成本往往高于试图阻止进入的现有企业选择的可变成本。

当然，现有企业阻止进入可能并不能获得收益——也就是说，很有可能 $\max R(Q,C) - N(C) > R[Q(C^d),C^d]/[r-\epsilon(C^d)]$，其中 C^d 表示阻止进入的技术，它是方程8的解。

稍后，在稍有不同的情况下，我将证明是否进入取决于市场上的企业数量是奇的还是偶的。在有着无穷多潜在进入者的情况下，我描述了一种混合策略均衡。这里也有一个具有这一形式的均衡：在每一期，都有某一企业以给定的概率进入，并替代现有的企业。同时，价格只是逐渐地降低。事实上，在通常的假设下——如果进入者定的价格恰恰小于市场通行价格，则进入者可以窃取整个市场——价格可以无限接近垄断价格。

19　如果在 T 之前没有潜在进入者进入，利润的贴现值就是 $[R(1-e^{-rT})/r] + e^{-rT}F_e$，第一个潜在进入者在 T 时到达的概率是 $\phi e^{-e\phi T}$。

如果

$$\epsilon = \{R[Q(C^m), C^m] + \phi F_e\}/(r + \phi)$$

则现有企业对于是否阻止进入是无差异的。现在令没有沉没的成本占总成本的比例为 ϕ（$=F_e/\epsilon$）。如果

$$\Phi \leqslant 1 - (r/\phi) \cdot [R/\epsilon r - 1]$$

则阻止进入。

由此出现了两个结果。首先，如果沉没成本很小（Φ 接近于1），只要潜在进入是有限的（Φ 是有穷的），价格就会设定在垄断价格水平上；并且不存在价格限制。一个垄断者由另一个垄断者替代，伴随着价格的无穷小的减少。第二，随着潜在进入者的数量增长到无穷，现有企业总是努力阻止进入。尽管这种反应会导致所谓的限制价格，但是价格并没有设定在竞争价格上，并且利润不等于零。只有进入成本（ϵ）趋近于零的时候，才能得到竞争结果。

4. 带有退出成本的串谋行为

在前一部分中，我们展示了进入者是如何将现有企业赶出市场的。进入不会导致竞争，只会导致退出；或者，进入的威胁会导致现有企业选择阻止进入的技术。

除了争夺和退出，现有企业也可以串谋。而进入者——通过选择技术——也可以促成串谋。假设现有企业选择了退出成本很高的技术，以使潜在进入者相信：现有企业不会退出。现在，进入者需要说服现有企业，对现有企业而言，合作比争夺好。进入者可以做到这一点，我将通过简单地扩展模型来证实这一点。假如存在三个时期，每一期都分成两个阶段。在第一阶段，每个企业同时做出进入—退出决策。在第二阶段，每一个企业都观察到其对手的退出—进入决策，同时决定价格。

证明的结构非常简单。如果进入者进入，在第三期就有三个均衡：现有企业退出，进入者退出，或者存在一个混合策略均衡。进入者承诺在第三期退出——让现有企业享有垄断利润——如果现有企业在第二阶段以合适的方式与进入者分享市场。换句话说，进入者通过声称在第三期退出来贿赂现有企业。如果现有企业不合作，进入者就会宣称采取混合策略，以一定概率不退出。如果现有企业合作，进入者会在第三期退出这一承诺是可信的，因为进入者在退出的时候，获得的利润至少和它继续留在市场上的利润一样多。同时，如果现有企业不合作，进入者就不退出（采取混合策略），这一威胁也是可信的。因为，如果进入者不退出，纳什均衡必然导致混合策略。给定

这一可信的承诺和可信的威胁，对现有企业而言合作是有利的。图4描述了这一均衡。

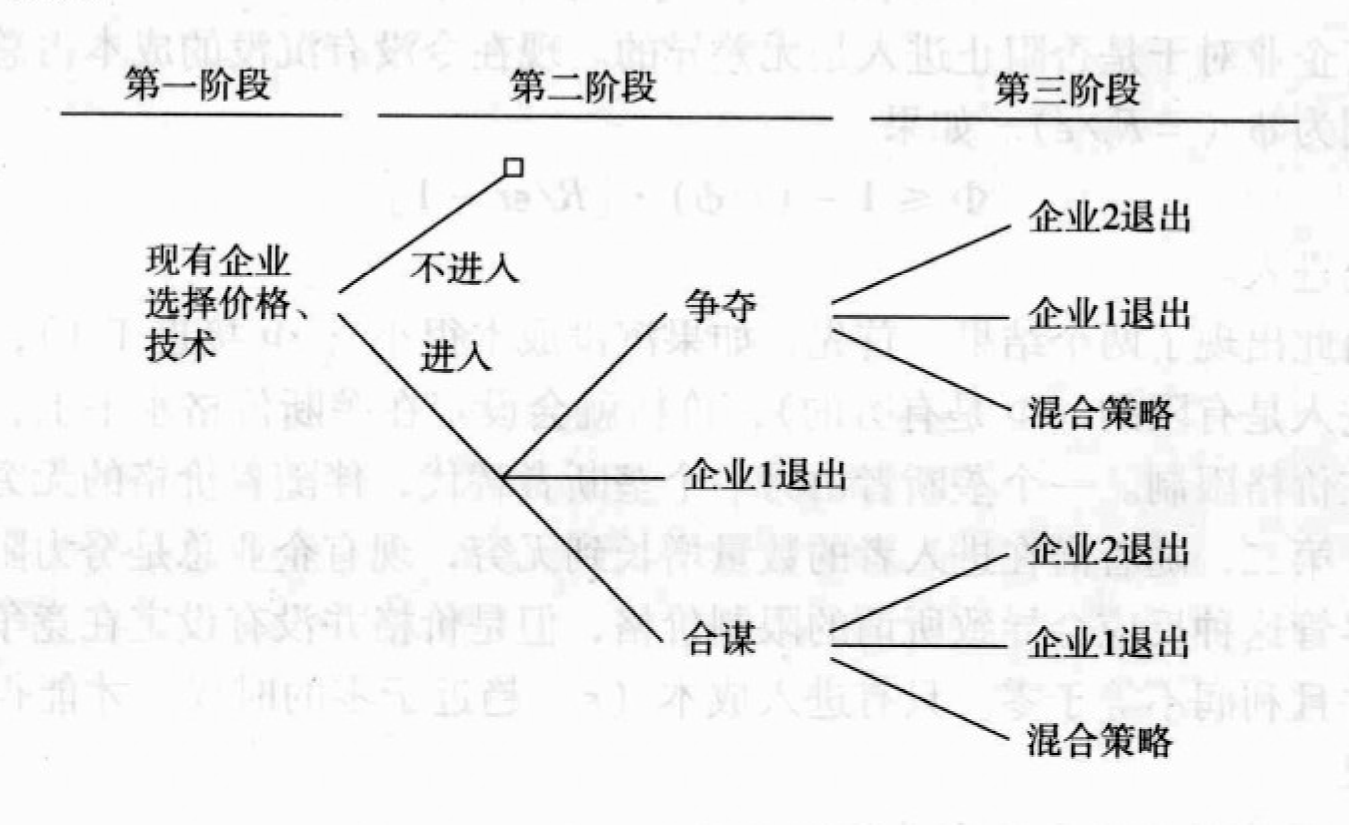


图4　串谋均衡

现在，我要详细讨论可以使串谋结果成为非合作均衡的一部分的情形。

假设存在一个时间折扣因子 $\delta = 1/(1+r)$，并且垄断利润 π_m 以 g 的速率增长。现有企业在退出的时候可以得到 F_I，进入者在退出的时候可以得到 F_e（F_I，F_e 既可以为正也可以为负）。进入者宣称，如果现有企业在第二阶段索取垄断价格，并且允许进入者分享 $(1-\alpha)$ 的市场，进入者就会在第三期退出。如果现有企业不合作，进入者就会在第三期采取非合作均衡混合策略。这使得进入者离开的概率为 β_e。当进入者同其对手都不退出的时候，进入者就会索取同生产的边际成本相等的价格。如果进入者的对手退出，进入者就会索取垄断价格。类似地，现有企业离开的概率为 β_I。如果现有企业不退出，它会采取和进入者相同的价格策略。退出获得的利润是 $F_i(i=I,e)$；进入者不退出的预期利润是 $\beta_I\pi_m$。因此需要通过选择 β_I 使得进入者在退出和不退出之间无差异：

$$\beta_I = F_e/\pi_m \tag{9}$$

通过选择 β_e 使得现有企业在退出和不退出之间无差异：

$$\beta_e = F_I/\pi_m \tag{10}$$

现有企业也宣布，如果进入者在第二期合作，现有企业在第三期会保持不变；如果进入者不合作，现有企业将会采用上文描述的混合策略。

合作的生存能力（*Viability of cooperation*）首先需要确认进入者的承诺——如果现有企业合作，进入者就会退出——是可信的。假设双方在第二期合作，在这种情况下现有企业不会退出。这时，只要 $F_e \geqslant 0$，进入者在第三期的最优行动就是退出（如果进入者不退出，它在第三期的利润就是零；当两个企业都在市场上的时候，就会出现伯川德均衡）。类似地，如果进入者总是会退出，只要 $F_I \leqslant \pi_m$，现有企业不退出就会更好。

接下来，我将证明如果不合作，就会采用混合策略这一威胁是可信的。以上描述的混合策略是一个纳什均衡，给定所有的企业都相信对方会采取混合均衡（如果在第二期不合作），企业也采取混合策略比较好。

现在考虑第二期。假设进入者进入了。如果现有企业合作，它的利润必然大于或者等于它不合作时的利润，也就是说

$$\alpha\pi_m + \delta g\pi_m \geqslant \pi_m + \delta F_I \tag{11}$$

如果现有企业决定欺骗，它就会在第二期索取低于进入者的价格，窃取全部的垄断利润；然后，在第三期退出，获得 F_I。

类似地，进入者最好不要欺骗，也就是说，

$$(1-\alpha)\pi_m + \delta F_e \geqslant \pi_m + \delta F_e \tag{12}$$

由此直接得到，如果要使进入者不欺骗，则 $\alpha = 0$。但是为了使现有企业愿意合作，我们要求（重写式 11）$F_I \leqslant (\delta g - 1)\ \pi_m/\delta$。

我们再次发现，进入并没有使消费者受益：价格仍然是垄断价格。唯一的变化就是垄断利润被分享了，但却造成了真正的社会成本。

在本文的附录 B 中，我证明了：如果扩展期数，第二期合作均衡（分享原则）的范围会增加。[20]

进入阻止。由此，我已经给出了在什么条件下，只要有进入，就会出现串谋均衡。如果上述条件满足，并且如果进入成本 ϵ 足够小，以至于 $\epsilon \leqslant \pi_m + \delta F_e$，就会出现进入。因此，为了阻止进入，现有企业必须选择一

---

20 事实上，更一般的理论是：随着期数变得无穷大，如果折扣因子是 1，那么本质上，所有的合作均衡都可能实现。如果存在类似这样的多重均衡，则佛克定理（Folk theorem）可以被扩展到有限期博弈。参见 Jean - Pierre Benoit and Vijay Krishna, "Finitely Repeated Games," *Econometrica*, Vol. 53 (July 1985), pp. 905 - 922。

实证经验表明，串谋行为可能比博弈论分析所建议的更容易达成。然而，这里我要构建的特定博弈论理论架构却受到了批评。因为将来的行为会受到过去的行为的影响，而过去的行为只有通过宣布的策略才可以影响到当前的状态。根据既往不咎这一原则，我们认为，在日期 t 开始的任何合理均衡，都应当只取决于 t 时的状态变量，或者对状态变量的信念。

种技术，使得合作是无利可图的，并且 F_I 很高。[21]

如果 $\delta = 1$ 并且 $g = 1$，为了不出现合作均衡，现有企业只需要使 $F_I > 0$。但是这么做，使得进入以一定概率出现。[22]因此，在这种情况下，仅有三种可能的结果：如果 F_I 是负的，进入者进入，双方串谋；进入威胁不对当前价格产生约束。如果 F_I 是正的，进入和退出可能出现，价格仍在垄断价格水平上；或者，偶尔会出现价格战。不考虑进入成本和沉没成本的大小，任何一种均衡都可能出现——随着沉没成本趋近于零，价格不一定收敛到竞争价格水平。即便有价格战，降低沉没成本对于预期价格的效应也是模糊的。随着沉没成本（作为总进入成本的一部分）趋向于零，仍然有一定概率使得第二时期的价格设定在垄断价格水平上。即便第二时期的价格在竞争价格水平上，仍然有一定概率使得价格在第三时期在垄断水平上。[23]

5. 内生的沉没成本

沉没成本（进入和退出成本）影响到事前竞争和事后竞争的特性。因为进入者和现有企业都知道这一点，他们在做投资决策的时候自然而然地考虑了这一点。换种说法，沉没成本代表了一种承诺。在本部分，我考虑一个简单的例子，其中现有企业还可以选择不同的技术，有些技术会产生沉没成本。

我想要说明两点。第一，许多关于竞争性的文献似乎都依赖于没有沉没成本的技术的存在。我要证明，事实上还需要其他的条件：没有沉没成本的话，就不可能有技术上有效的技术。第二，为了阻碍进入，现有企业付出的沉没成本可能比经济上有效率的沉没成本要大；因此，潜在竞争可能再次导致帕累托次优均衡。[24]

21　尽管不存在合作均衡，但可能存在一些其他的完美均衡——有些导致进入，有些不会。例如，现有企业宣称它的策略是：在任何有进入的时期，它都会索取竞争价格；在接下来的各期，它都将采取上文描述的混合策略。这样，只要 $\epsilon \geqslant \delta F_e$，也就是说，只要有一些不可偿付的（以折现价值表示）沉没成本，就存在一个没有进入的完美均衡。在这个均衡中，现有企业会索取垄断价格：进入的威胁是无效的。

22　存在一个完美均衡：进入者进入，现有企业退出，价格设定在垄断价格水平上。还有另一个完美均衡：进入者以一定概率进入，现有企业以一定概率退出。当两个企业都在市场上的时候，价格是伯川德价格；当只有一个企业在市场上的时候，价格被设定在垄断价格水平上。通过选择进入和退出的概率来保证：对在位者而言，预期利润为零；对进入者而言，预期利润等于进入成本。

23　事实上，如果进入发生了，随着沉没成本的减少，以下这一情况就越来越有可能：两个企业之一，甚至很有可能两个企业全都会在第三期退出。

24　这一分析是从不确定性中提炼出来的。对于未来需求和技术的不确定性给出了另一个原因，解释为什么沉没成本可能不是经济上有效的（从社会角度）。作为阻止进入的一种方法，除了我们这里所讨论的一些成本之外，沉没成本可能还有其他成本。

考虑这样一种情况：有两种技术供一个企业选择，其中一种技术有沉没成本，另外一种技术没有沉没成本。一个自然的假设是，通过使投资不可偿付（沉没），企业在可变成本上可以有所获益。如果没有不确定性，企业会选择沉没成本更有效率的技术（如果产出足够大）。只要使资本还可以在其他地方或者其他行业使用，需要付出一定成本，适当的均衡就一定会产生沉没成本。一个可能和事实相对的假设——存在一种没有沉没成本的技术——确实有着重要的意义：它限制了企业运用垄断力量的程度。它无法索取过高的价格，使得没有沉没成本的企业可以进入，采取比该价格稍低的价格，并赚取利润。而这反过来又有重要的含义：现有企业可以选择一种技术来阻止进入，甚至没有沉没成本的企业也会被阻止进入。因此，假设有两种技术供企业选择。其中一种技术的沉没成本比较高，但是在产量到达某个比较高的产量 Q_{max} 之前（见图5），生产的边际成本为零。另外一种技术沉没成本较小，但在产量到达 Q_2 之前，生产的边际成本是正的。假设在垄断的产出水平上，第二种技术比第一种技术有效率（即便对于在垄断产出水平上，边际成本为零的技术而言，我们也假设这一点是成立的）。现在考虑第三种没有沉没成本的技术——竞争性技术。如果企业可以赚取利润，或者至少收支平衡，企业就会采用这种技术进入。如果现有企业选择了第一种技术，在伯川德均衡中，假如进入者进入市场，现有企业就会生产 Q_{max}。而如果它选择第二种技术，它就会生产 Q_2。分别将潜在进入者在两种情形下面临的剩余需求曲线记为 $D'D'$ 和 $D''D''$，并且假设 $D''D''$ 和平均成本曲线相交，而平均成本曲线总是高于 $D'D'$（见图5）。因此，通过选择无效率的技术（带有更大沉没成本的），企业甚至可以阻止没有沉没成本的企业进入。这么做是否利

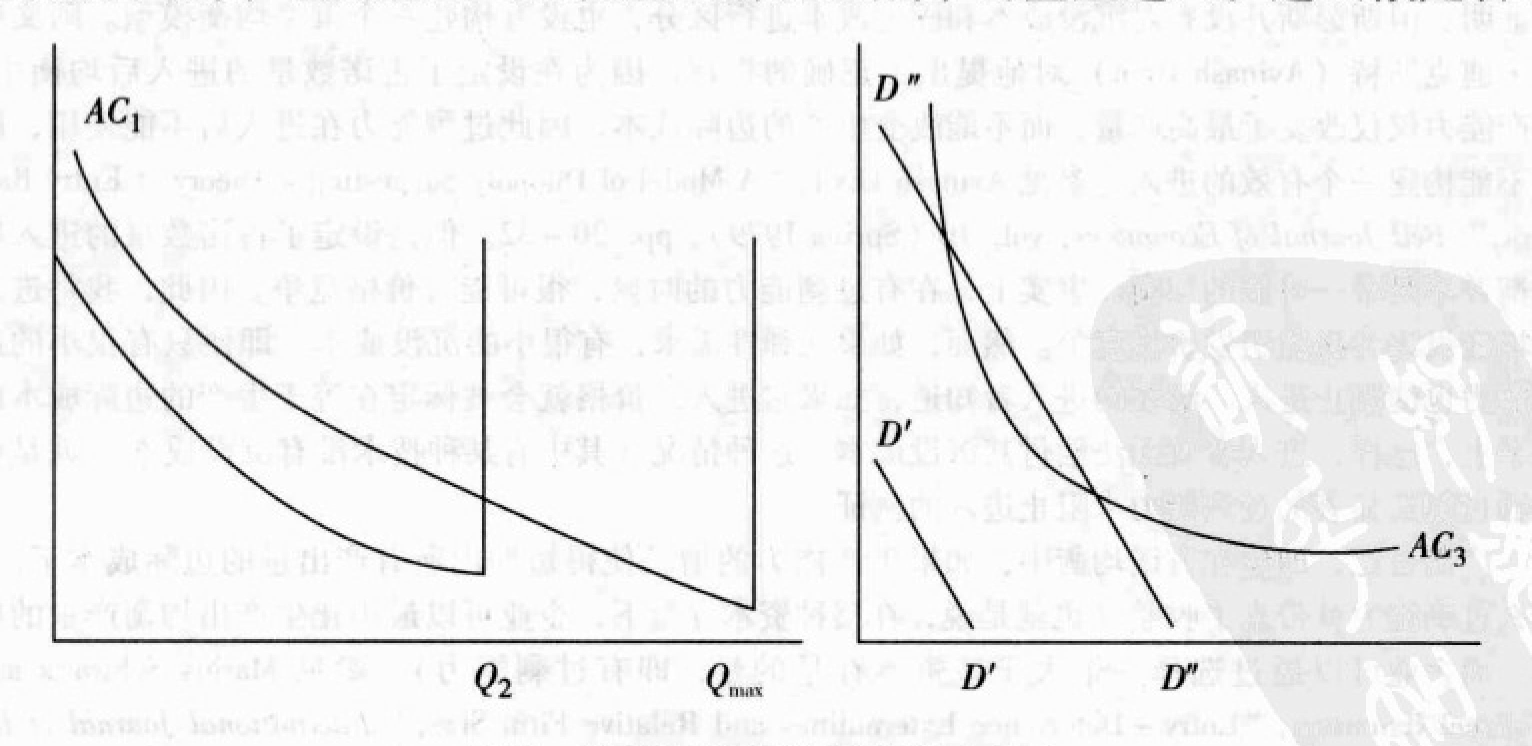


图5　使用无效率技术来阻止进入

润更丰厚，取决于采用无效率技术带来的垄断利润的大小，与采用第二种技术产生的竞争均衡中均衡价格带来的利润的大小的比较。两种情况都是可能的。[25]

如果企业采取无效率的方法来阻止进入，它们自然而然地会追问自己，采用其他的策略是否能同样有效地阻止进入，哪一种策略实施的成本最小。我们可以想像，企业可以同第三方签订很多种合约，约定如果发生了进入，第三方不进行报复，则企业将付给第三方一大笔固定资金。这样的合约保证了在进入后的博弈中会有激烈竞争是可信的，因此可以阻止进入。并且，因为我在分析中假设了不随机，所以进入不会发生，这些合约也不会产生成本。然而，如果进入发生了，在第二期，第三方和现有企业都有动机重新签订合约。因为第三方意识到，如果不重新签订合约，它将什么都得不到，而如果重新签订合约，它至少可以得到点什么。最初的现有企业也可以通过重新签约而获利。了解了这一点，进入者可能就不会相信这些合约是可信的。

还有两种说法可以解释为什么没有出现这样的合约。沉没成本可能足够重要，以至于它们可以成为无成本的进入壁垒，因此合约是没有必要的。另

25　均衡价格是下式的解 $D(Q_2+Q_c)=A(Q_c)$，其中 $A(Q)$ 是竞争性技术的平均成本曲线，Q_c 是采用该技术的企业的产量。因此，采用技术 2 的现有企业两期的利润是 $R[Q(C_2),\ C_2]+[A(Q_c)-C_2]\ Q_2-\epsilon_2$（不考虑折现因素），采用阻止进入的技术带来的利润是 $2R\ [Q(C_1),\ C_1]-\epsilon_1$，其中 C_i 是采用技术 i 的边际成本，ϵ_i 是沉没成本，$R(Q,\ C_i)$ 是采用技术 i，可变成本为 C_i，产出量为 Q 的垄断者获得的运营利润，$Q(C)$ 是利润最大化的产出水平。

我认为，这是对由迈克·斯彭斯（Michael Spence）提出的用过剩能力作为阻止进入的手段的正确证明。但斯彭斯并没有对沉没成本和固定成本进行区分，也没有构建一个策略均衡模型。阿文纳什·迪克斯特（Avinash Dixit）对他提出了正确的批评，因为在设定了古诺数量的进入后均衡中，生产能力仅仅改变了最高产量，而不能改变生产的边际成本，因此过剩能力在进入后不能使用，因此不能构建一个有效的进入。参见 Avinash Dixit, "A Model of Duopoly Suggesting a Theory of Entry Barriers," *Bell Journal of Economics*, vol. 10 (Spring 1979), pp. 20 – 32。但是设定了古诺数量的进入后均衡并不是唯一可能的均衡。事实上，在有过剩能力的时候，很可能有价格竞争。因此，我将进入后均衡设定为必然引发价格竞争。然而，如果无弹性需求，有很小的沉没成本，即便只有很小的过剩能力也能阻止进入。潜在的进入者知道，如果它进入，价格就会被标定在等于生产的边际成本的水平上，这样，进入者就无法偿付其沉没成本。这种情况（其中有某种技术没有沉没成本）就是可以通过创造显著的过剩能力来阻止进入的例证。

广而言之，即便在古诺均衡中，如果生产能力的增长使得短期内所有产出量的边际成本下降，那么过剩能力就带来了收益（也就是说，在某种资本存量下，企业可以最小化生产出均衡产量的成本，而企业可以通过选择一个大于该资本存量的量，即有过剩能力）。参见 Marius Schwartz and Michacel Baumann, "Entry – Deterrence Externalities and Relative Firm Size," *International Journal of Industrial Organization* (forthcoming)。

外可能还有一些重要的随机因素。

一般的观点依然成立：现有企业对技术的选择会受到进入威胁的影响。它可能会选择一种能够阻止进入的技术。它可能选择一种技术，使得一旦两家企业激烈地竞争，则企业的进入就是无利可图的。现有企业可能选择一种技术，使得进入者相信，现有企业不会退出。或者，它可以选择一种技术，使得进入者相信它不会合作。现有企业会选择沉没一定的成本：竞争学说要求不能有带有沉没成本的技术，这是一个很难受认可的假设。一般而言，为应对竞争威胁而采取的技术不会是有效率的技术。

6. 租的消散

我们从上一部分学到了一个一般的教训：潜在竞争并不足以保证零利润。

行业里的第一个企业有一定优势，这是不能被后来者完全消除的。毫不惊讶的是，为了成为行业里的第一个现有企业，企业间会展开竞争，这一竞争也许会也许不会完全消除潜在租。

如果有很多竞争者，他们地位相同，那么争相成为第一个企业的竞争，就可以被看作是一场专利竞争。在某些情况下，这并不会导致租值消散。[26]甚至对于取得目前在商业上不可行的产品的专利的企业而言，这也是有利可图的。在这种情况下，早期的研究从社会角度来讲也不是值得要的，应当直到该商品在商业上可行时才适用。早期的研究是租值消散的一种形式。[27]在干中学的情况下，企业在价格低于边际成本的时候就开始生产（赚得负的利润流）是有利可图的。[28]

然而，在开始的时候，企业很少处于相同的位置。处于优势地位的企业只需要在研发上花费有限的资金，就可以预先阻止其竞争对手。之后我们就会看到这一点。然而，首先，我需要强调一些针对我的基本结论的反对意见。

二、反对意见

在前一部分，我们提出即便只有很小的沉没成本，潜在竞争也不足以保

26 Drew Fudenberg and others, "Preemption, Leapfrogging, and Competition in Patent Races," *European Economic Review*, vol. 22 (June 1983), pp. 3 -31.

27 Partha Dasgupta, Richard J. Gilbert and Joseph E. Stiglitz, "Invention and Innovation under Alternative Market Structures: The Case of Natural Resources," *Review of Economic Studies*, Vol. 69 (October 1982), pp. 567 -582.

28 Partha Dasgupta and Joseph E. Stiglitz, "Learning by Doing, Market Structure, and Industrial and Trade Policies," *Oxford Economic Papers* (forthcoming).

证经济效率或者保证有效生产的收益会被传递给消费者——也就是说，利润将被降为零。另外，甚至进入本身也不能确保有效竞争，因为现有企业可能会容忍进入，并且与之串谋。

那些支持这一观点——潜在竞争可以确保经济效率和零利润——的人对我的分析提出了反对意见，意见大体分为两类：一类是认为我假设的竞争过于激烈，还有一类是认为我没有考虑到竞争的所有可能性（也就是说，我假设的竞争不足够激烈）。

如果竞争不如生产同质产品的伯川德竞争激烈，那么在进入之后，企业会获利，而这些利润会吸引更多的进入。当然，出于同样的原因，无论是潜在竞争，还是有限的真实竞争，都不足以保证零利润或者经济效率。政府介入是有理由的。

有限竞争可以采取三种形式：调整的滞后，不完全可替代性以及古诺（而不是伯川德）反应。

1. *滞后*

如果现有企业不能在进入者进入的时候，立刻降低其价格，那么潜在竞争——进入威胁——可能会影响企业的价格决策。如果企业要价过高，进入者就会进入，并且在现有企业可以做出反应之前抢走其顾客。因为进入之后的竞争是有限的，就有进入的动机；这正是使得潜在竞争可以发挥约束作用的原因。

这些反应的滞后，本质上是沉没成本的证明。如果企业在给定期限内不能改变价格，那么定价决策就好像一个沉没的，暂时无法偿付的投资决策。如果说定价决策的调整成本，高于生产决策的调整成本，因此要改变前者所花费的时间比改变后者要花费的时间长，这是否是正确的？某个著名的例子宣称，沉没成本在航空业中相对而言并不重要，在进入之后价格会迅速下跌，而退出之后价格会迅速升高。很明显，航空公司（以及潜在进入者）并不认为进入前的价格在进入后也必须适用。事实上，我们认为当现有的航空公司未能做出降价反应的时候，它们只是因为害怕受到反托拉斯起诉。因此，是政府政策限制了企业，使之无法做出反应。但是，对这一研究的反对观点会问，针对这样的价格反应，政府政策应当怎样呢？

因为我相信，定价决策中的滞后反应（尤其是关系到进入和退出的），通常都比生产决策中的滞后反应短。我已经集中关注了那些就相对调整速度做出判断的模型。但是在本部分，我将探讨较长的价格反应的后果。这些对于反应滞后的限制，被想像为限制事后竞争（以及增加事前竞争）的一种方式，

但是它们作为限制事后竞争的方式，既是难以置信的，也是没有效率的。

为了检验价格滞后的效应，我稍微地调整了先前的模型，允许企业不能通过降低价格迅速地对进入做出反应。对于价格刚性假说，有两种版本的说法：在我们即将探讨的那个版本中，企业必须提前 L 期宣布价格调整，但是它可以随心决定价格变化的频率。[29]在第二个版本中（稍后讨论），企业每 L 期才可以调整自己的价格。在第一个版本中，有两种可能的均衡模式：容忍进入和阻止进入。如果把模型改为连续时间，就非常容易看到这一点了。在时点 0 和时点 1 之间，现有企业有着安全的垄断地位；当进入发生的时候，如果它发生了，必然是在时点 1 或者在时点 1 之后。[30]

进入阻止。在 L 期之后市场就退化为传统的伯川德均衡，并且价格等于边际成本。因此，进入者只有在 L 期内现有企业价格固定在 q_1 时才可以赚得利润。令 $\pi(q)$ 表示索取价格 q 时（忽略沉没成本）可以赚得的利润（此时，对手索取的价格高于 q，或者没有对手）。因此，进入者的利润的贴现值是：

$$[\pi(q)\cdot(1-e^{-rL})]/r = V(q,L,r)$$

如果现有企业设定其价格，满足 $V(q,\ L,\ r)\ =\epsilon$，就可以阻止进入。因此

$$q_1 = \min(q_m, q)$$

其中，q_m 是垄断价格，也就是说 $\pi'(q_m)\ =0$。对于满足 $q_1=\min(q_m,\ q)$ L 的任何值，存在 ϵ 的临界值 ϵ^*，满足 $\epsilon>\epsilon^*$，$q_1=q_m$。正是（ϵ，L）组合满足 $V(q_m,\ L,\ r)\ =\epsilon$。

对于很大的 ϵ，竞争威胁对价格完全没有影响。对于 $\epsilon<\epsilon^*$，竞争威胁使得价格降低到垄断价格以下，但不会降到竞争价格水平（生产的边际成本）。只有当 L 接近无穷的时候，价格才会达到使进入者收支平衡的水平（如果进入者永远都是市场上唯一的生产者）（因此，在图 6 的Ⅱ区，竞争的威胁对现有企业完全没有任何影响；在Ⅰ区，竞争威胁有一定影响，但影响有限。需要注意，对于很短的反应滞后，即便有小的进入成本，现有企业也可以成为垄断者）。

竞争威胁对于消除利润也是无效的。如果 $\pi_1(q)\ \equiv\pi_2(q)\ =\pi(q)$，企业 1（现有企业）的净利润是

$$\pi(q/r-\epsilon_1) = \epsilon_2/(1-e^{-rL}) - \epsilon_1$$

29 在 $T=1$ 之前不能做出第一个宣告。

30 为了简便，允许现有企业在进入者确实进入之后，宣布降低价格；很明显，现有企业如果可以预期到进入的日期，有效的反应滞后期就会减短。

其中 ϵ_i 是第 i 个企业的进入成本。如果 $\epsilon_1=\epsilon_2=\epsilon$，随着 $\epsilon\to 0$ 或者 $rL\to\infty$，利润的贴现值是 $\epsilon e^{-rL}/(1-e^{-rL})\to 0$。

容忍进入。在阻止进入之外，现有企业还可以选择容忍进入。现有企业宣布以下策略：只要进入者索要价格 q_m，现有企业也将索要价格 q_m；如果进入者索取一个低于 q_m 的价格，则进入者会恢复到伯川德策略，索要比对手更低的价格。如果进入者也宣布同样的策略，则存在一个完美均衡。如果双方合作，利润的贴现值就是 $\pi(q_m)/2r$。因为当双方都采取伯川德策略的时候，利润是零，因此双方都能从合作中获得好处。

现有企业是会选择容忍还是选择阻止进入，取决于在两种情况下利润的相对价值，也就是说，取决于 $\epsilon_2/(1-e^{-rL})$ 是否大于 $\pi(q_m)/2r$。图 6 中的虚线恰好表示使企业觉得无差异的 (ϵ_2, L) 轨迹。因此，在进入成本很小的时候，价格回到垄断价格水平。现有企业容忍了潜在竞争者，两家企业进行串谋，瓜分市场。[31]

价格滞后的其他解释。在对于反应滞后的其他解释中，现有企业每过 L 期才可以改变一次价格。考虑到对称性，进入者也是每过 L 期才可以改变一次价格。进入者必须决定什么时候进入。为了做出决定，它必须考虑到这一事实：现有企业会如何反应，取决于进入者的价格选择以及进入日期。现有企业可以选择争夺或者合作，也就是说和进入者分享市场。

假设在现有企业设定价格后 L^* 期，进入者进入市场。我要首先证明，此时容忍是不可能的。假设现有企业将价格降为恰恰比进入者的价格低一点，借此来反击进入者，而在轮到进入者调整价格的时候，它也将价格降为恰恰比现有企业的价格低一点。因此，现有企业的利润贴现值是：

$$\pi(q^*)(1-e^{-rL^*})/r(1-e^{-rL}) \tag{13}$$

如果现有企业采取合作策略，索取和进入者同样的价格，其利润的贴现值就是：

$$\pi(q^*)/2r \tag{14}$$

它必须设定 (L^*, q^*) 来诱导合作，而不是竞争。另外，当进入者有机会重新设定价格的时候，合作比争夺更有利。如果争夺，进入者的利润的贴现

31　当有很多潜在进入者的时候，如果随时可以进入，则完美均衡可能会导致没有进入。假设一个企业在某个特定日期进入市场，并且此后 L 期内不能改变它的价格。另外一个潜在进入者就会发现稍后进入是有利的——它可以索取比前一个进入者低 ϵ 的价格，从而窃取市场。因此，第一个进入者获得的利润会过于小，以至于不能偿付其进入的固定成本。我们再一次得到了看似矛盾的结果：竞争越多，导致的竞争越不激烈。

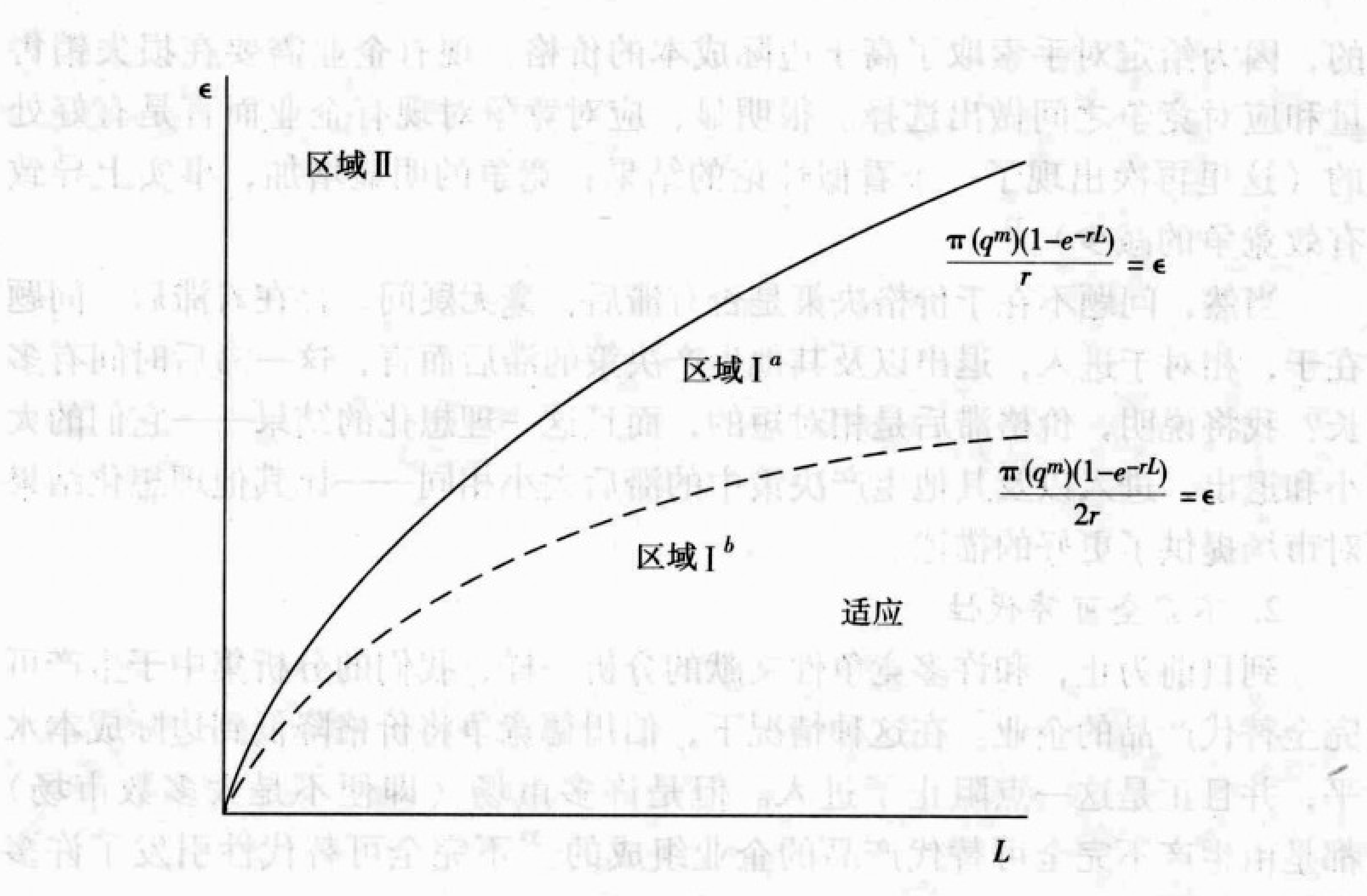


图6　反应滞后和进入

值就是：

$$\pi(q^*)[1-e^{-r(L-L*)}]/r(1-e^{-rL}) \tag{15}$$

而如果它合作，其利润贴现值就如式（14）所示。因此，为维持合作，必须保证：

$$0.5(1+k)\leqslant\Gamma\leqslant 2k/(1+k)$$

其中 $k=e^{-rL}$，$\Gamma=e^{-rL^*}$。但是由于 $k\leqslant 1$，这一不等式不能得到满足。所以，长期的合作是不可行的。因此，对于进入者而言，恰好在进入者设定价格之后进入是有利的。并且，每当进入者要重设价格的时候，使价格恰恰低于在位者的价格，对于进入者而言是有利的。因此，即便有价格滞后，市场在任何时候都是由一个主导出售者刻画的。尽管价格下降，调整滞后意味着它将下降得很慢。为实用的目的，价格将大致停留在垄断价格附近。

价格反应滞后的合理性。尽管我已经指出，价格反应滞后的存在并不足以重建竞争学说的有效性，我还将进一步说明显著的价格反应滞后都是不太合理的（与之相对应的是，生产滞后可能是显著的）。事实上，对于现有企业而言，宣布一项政策来应对竞争，索取一个恰恰低于对手的价格，使价格降到生产的边际成本上很容易。企业可以并且确实将这些政策写入了销售合约。这样的策略确保了没有进入。并且，这一策略是可信

的，因为给定对手索取了高于边际成本的价格，现有企业需要在损失销售量和应对竞争之间做出选择。很明显，应对竞争对现有企业而言是有好处的（这里再次出现了一个看似悖论的结果：竞争的明显增加，事实上导致有效竞争的减少）。[32]

当然，问题不在于价格决策是否有滞后：毫无疑问，存在着滞后。问题在于，相对于进入，退出以及其他生产决策的滞后而言，这一滞后时间有多长？我将说明，价格滞后是相对短的，而且这一理想化的结果——它们的大小和退出，进入以及其他生产决策中的滞后大小相同——比其他理想化结果对市场提供了更好的描述。

2. 不完全可替代性

到目前为止，和许多竞争性文献的分析一样，我们的分析集中于生产可完全替代产品的企业。在这种情况下，伯川德竞争将价格降低到边际成本水平，并且正是这一点阻止了进入。但是许多市场（即便不是大多数市场）都是由生产不完全可替代产品的企业组成的。[33]不完全可替代性引发了许多有趣的问题。

*不完全可替代性和进入阻止。*不完全可替代性的一个后果是，事后竞争将不那么激烈，因此进入更加容易。比如，假设需求函数族由具有 $a(q_2/q_1)\ I/q_1$ 形式的间接效用函数推导出来，其中 I 是个人的收入，在 $q_2=q_1$ 的时候，$a'q_2/aq_1=0.5$。则可以证明，在进入后的博弈中，均衡价格将是

$$q_1 = q_2 = C(\sigma+1)/(\sigma-1)$$

其中 σ 表示两种商品之间的替代弹性。很清楚，假如 σ 小于无穷大，则价格将高于生产的边际成本。图 7 显示了 ϵ 的临界值（σ 的函数），在临界值

32　斯蒂文·萨勒普（Steven C. Salop）较早地注意到了采取应对竞争的策略（或者最惠国（most favored nation））的后果。参见 Steven C. Salop, "Practices That (Credibly) Facilitate Oligopoly Coordination," in Joseph E. Stiglitz and G. Frank Mathewson, eds., *New Developments in the Analysis of Market Structure* (MIT Press, 1986), pp. 265 - 290。托马斯·库帕（Thomas E. Cooper）进一步分析了这些后果，请参见：Thomas E. Cooper, "Facilitating Practices and Most - Favored Customer Pricing" (Ph. D. dissertation, Princeton University, 1984)。

33　如果这是真实的，正确的经济模型显然应该是不完美的或者垄断竞争，而不是完美的、竞争的阿罗—德布鲁模型。垄断竞争经济和完全竞争经济之间有重大区别，其中包括：一般来说，福利经济学基本定理不适用于它们。要了解对这些模型的近期研究，请参见 Joseph E. Stiglitz, "Towards a More General Theory of Monopolistic Competition," in M. H. Preston and Richard E. Quandt, eds., *Prices, Competition and Equilibrium* (Deddington, England: Philip Allan, 1986), pp. 22 - 69。

之上不会有进入。

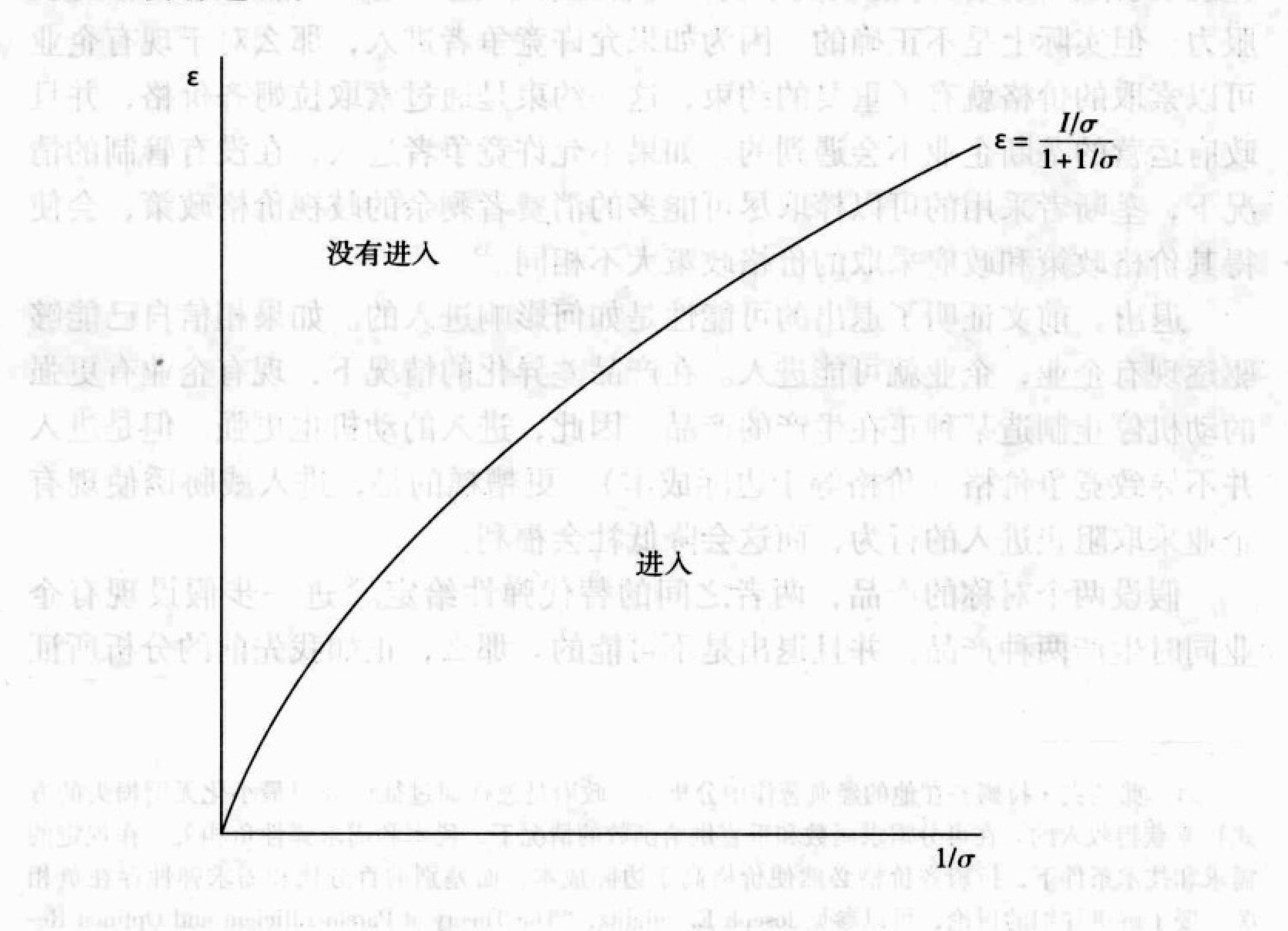


图7　进入和不完全可替代性

多产品企业和拉姆齐价格。当我们把对生产同质产品的市场的关注转移到生产相关产品的市场上的时候，我们需要研究一整套市场结构。行业内一个企业可以生产所有的产品。第二个企业进入的时候，既可以就某一产品与之竞争，也可以就整套产品与之竞争。这是那些负责公司战略的经理们所面临的问题——无论是那些考虑进入市场的经理，还是那些试图保护自己的企业采取阻止进入的经理。

产品差异化的一个原因是，生产各种产品的固定成本。因为非凸性是产品差异化市场分析的核心，竞争学说对于分析这样的市场是怎样运行的尤其有用。并且，支持竞争学说的人，事实上已经考虑了这样的行业。他们认为即便技术使得只有一家企业生产，竞争性也确保了它将在零利润下运营。它索取拉姆齐价格，用收入来偿付固定成本，相应地，它完全像政府企业

（受到约束，使其恰好能收支平衡）那样运作。[34]这一论证虽然看起来很有说服力，但实际上是不正确的。因为如果允许竞争者进入，那么对于现有企业可以索取的价格就有了重要的约束，这一约束是通过索取拉姆齐价格，并且政府运营的垄断企业不会遇到的。如果不允许竞争者进入，在没有管制的情况下，垄断者采用的可以榨取尽可能多的消费者剩余的歧视价格政策，会使得其价格政策和政府采取的价格政策大不相同。[35]

退出。前文证明了退出的可能性是如何影响进入的。如果相信自己能够驱逐现有企业，企业就可能进入。在产品差异化的情况下，现有企业有更强的动机停止制造某种正在生产的产品。因此，进入的动机也更强。但是进入并不导致竞争价格（价格等于边际成本）。更糟糕的是，进入威胁诱使现有企业采取阻止进入的行为，而这会降低社会福利。

假设两个对称的产品，两者之间的替代弹性给定。[36]进一步假设现有企业同时生产两种产品，并且退出是不可能的。那么，正如我先前的分析所证

---

34 弗兰克·拉姆齐在他的经典著作中分析了，政府是怎样通过征税（以最小化无谓损失的方式）来获得收入的。在可分需求函数和垂直供给函数的情况下，税率和需求弹性负相关。在规定的需求和技术条件下，拉姆齐价格必然使价格高于边际成本，而差别的百分比和需求弹性存在负相关。要了解更详细的讨论，可以参见 Joseph E. Stiglitz, "The Theory of Pareto-Efficient and Optimal Redistributive Taxation," in Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, vol. 2 (North-Holland, 1987), chap. 15；或者 Anthony B. Atkinson and Joseph E. Stiglitz, *Lectures on Public Economics* (McGraw-Hill, 1980)。

35 要了解关于这一点的更全面的讨论，请参见 Sappington and Stiglitz, "Information and Regulation"。这些研究也注意到了更多的问题，无论是理论上的还是实证上的，这些问题是在互相依赖并且非常普遍的情况下，由对行业边界的定义引起的。如果采用比较宽泛的定义，把市场经济看作分散分配机制的看法就是受到损害的。如果采用比较狭隘的定义，商品集（对其征收的税可以用来提供收入，为固定成本提供资金）是不必受到限制的。

另外需要明晰的是，如果企业可以采用非线性价格，即便只能索要一笔固定的服务费，最优价格集也和拉姆齐—波提奥克斯价格（Ramsey-Boiteaux prices）没有联系。请参见 Anthony B. Atkinson and Joseph E. Stiglitz, "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economies*, vol. 6 (July-August 1976), pp. 55-75。

以下这个流行的观点是很让人吃惊的：在非凸性和自由进入的情况下，市场均衡将是帕累托有效的。早在 Baumol, Panzer and Willig 的研究 *Contestable Markets* 之前，Avinash K. Dixit and Joseph E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, vol. 67 (June 1977), pp. 297-308，已经证明了，只有在非常紧的条件下（紧到难以满足的程度），零利润条件才能确保受限的帕累托最优（也就是说，政府无法对企业提供"一揽子"补助）（所有的企业都必须面对产品的不变需求弹性曲线，需求弹性相同）。

36 肯尼斯·贾得（Kenneth Judd）给出了这个模型的一个变体，请参见 Kenneth Judd, "Credible Spatial Preemption," *Rand Journal of Economics*, vol. 16 (Summer 1985), pp. 153-166。

明的，如果有 ϵ 的沉没成本，就不会有进入（如果有伯川德竞争）。现在，假设现有企业在退出的时候，从某一条生产线可以获得 F_1。令 π^d 表示企业从某条生产线退出而可以获得的利润，π^b 表示不退出而可以获得的利润。[37] 如果 $\pi^d + F_1 > \pi^b$，那么现有企业退出某一条生产线会获利。但是，如果这一方程满足，只要 $\epsilon < \pi^d$，就会有进入。现有企业知道这一点，并且有动机选择退出成本很大的技术；这些技术可以作为进入阻碍。即便生产成本高一些，企业也会选择这样的技术。因此，企业由可以承诺不退出而获得先动优势。[38]

3. 纳什—数量均衡（古诺）

到目前为止，我在分析中假设，两个企业在进入后博弈中，采取伯川德策略。假设没有任何生产能力约束，并且采用规模报酬不变的技术，这似乎是合理的。然而，我要简要探讨一下，两家企业在进入后博弈中，采取数量设定策略的后果。

古诺竞争没有伯川德竞争那么激烈。这意味着，进入后的利润会比伯川德竞争的利润大，这反过来意味着，沉没成本作为进入壁垒就更加无效。

比如，假设行业需求曲线的弹性不变：

$$Q = kc^{\mu-1}p^{-\mu}$$

那么进入后的均衡价格就是：

$$p = c/[1 - (1/2\mu)]$$

而进入后的利润是：

$$\pi = k[1 - (1/2\mu)]^{\mu-1}/2\mu$$

对于 μ 的任何值，都存在一个 ϵ 的临界值，在它以上就不会发生进入。临界值对 μ 是单调的（对于相关的 μ 值，也就是说，对于 $\mu \geqslant 0.5$），同时随着弹性趋向于无穷，ϵ 趋向于零。对于有限需求弹性，存在有限的沉没成本水平，在其之下就会产生进入。但是进入后的价格并不是竞争价格：即便是真实的竞争也不足以确保价格等于平均成本，更不用说边际成本了。

当存在大量的潜在进入者的时候，利润（扣除掉进入成本）减少为零；

37 在存在直接的（伯川德）竞争的生产线，企业无法获得任何利润，并且该生产线的低价降低了其他生产线的利润。

38 然而，现有企业可能会面临一个悖论。先前，我证明了较低价值的 F_l（高退出成本）可能会使进入企业相信：如果它进入了，现有企业会合作（因为退出没有好处），并与之分享垄断利润。可能不存在任何价值的 F_l 可以阻碍进入。

用均衡中的企业数量，n^*，可以解出 $\epsilon = k\ [1 - (1/n\mu)]^{\mu-1}/n\mu$。仅仅当 ϵ 趋向于零时，n 趋于无穷，并且价格收敛到边际成本水平。另外，如果进入自由，尽管一些竞争收益以低价的形式转移给了消费者，还有一些竞争收益却因为过多的进入花销而消散了。由于竞争而浪费掉的总资源是 $(n^* - 1)\epsilon$。随着 ϵ 趋向于零，这个值趋向于 k/μ。[39] 作为对商品的最优总花销的一部分，这一浪费达到了 $1/\mu$；也就是说，对于合理的需求弹性值，浪费掉的部分是不可忽略的。尽管利润是零，经济却远不是有效率的。

这里引发一个和前文伯川德竞争的例子中相同的问题：如果企业不能改变它在 L 期内的生产水平，会发生什么情况呢？在之前的例子中，我注意到，如果现有企业通过试图阻止进入，来回应潜在进入的威胁，潜在竞争会在进入前的时期降低价格。如果企业对竞争威胁的回应是容忍竞争，潜在竞争就会对进入前的价格没有影响。现在，如果现有企业意欲容忍进入，潜在竞争就可能在进入前的时期提升价格。因为，如果企业预期到了进入，则进入后时期的最优产出就会比只有一个生产者时的产出低；它将进入前的产出设定在垄断水平和双寡头水平之间。

4. 契约

在这之前的讨论，考虑了根据过去经济发展情形分析的竞争没有比之前预想的竞争程度激烈（因此，根据经济预期的变化分析的竞争可能比预想的更激烈）的原因。现在，我将考虑为何即便存在有效的事后竞争，潜在竞争也是有效率的。我将考虑两个原因。关于先前结果的一个批评是：我忽略了一个重要的竞争维度——契约，另一个批评是：我一定暗中假设了有限的潜在竞争。对这两个问题进行更深入的检验，我们发现，潜在竞争可能会比前一部分——“沉没成本和潜在竞争”——的分析结果更无效率，而不是更有效率。

那些认为潜在竞争会确保经济效率的人会认为：在有沉没的生产成本的情况下，在签订契约的阶段就有竞争。也就是说，在成本沉没之前，企业就在争夺顾客。但是，这一行为也涉及一些沉没成本。为建立一个协商合约的基础，必然涉及了解市场，树立一定的声誉等行为。这样的沉没成本可能不大，但是它们确实存在。[40] 并且，前一部分的论断认为：只需要很少的沉没

39 也就是说，$n\epsilon = k\ [1 - (1/n\mu)]^{\mu-1}/u$，它收敛到 k/μ。

40 当然，事实上，可能还有其他原因导致契约竞争是有限的。尤其是在某些研发市场上，实质上不可能设计出一种契约，能够完全细化将要出售的新产品的特质，以及出售的时间和价格。

成本，潜在竞争就可以产生有限的效率。

事实上，假如签订契约有一定的沉没成本，那么可以签订并执行长期契约的经济，可能比没有契约的经济的竞争性更小。成本必然会沉没，这一事实会分解市场，将一个看似很大的市场（拥有众多供给者和顾客）转变为许多小的市场，每一个小市场都有有限的供给者。换一种说法，当有契约的时候，现有的企业会锁定顾客，使得难以成功进入。我们通过一个高度程式化的例子来说明这一点：在均衡的时候，几乎所有的顾客都被锁定了，潜在竞争极其无效。[41]

时间是离散的。在每一个日期，都有 N 个个人诞生，每个人正好活 T 年——这同样也是机器的使用年限。造一台机器的成本是 F，且机器可以以零边际成本制造至多 M 个产品。每个个人在一生中的每一个日期都正好购买一个产品。个人在每一个日期对产品的保留价格是 u（>0）。如果$M=N$，则这个例子就极大地简化了。这样，效率就要求每一个日期建造一台机器，t 日生产的机器服务于 t 日出生的一代人，直到机器和这代人都损耗（死去）了（在$t+T$日）。事实上，如果T很大，可能会有很多的机器和企业，那么我们就倾向于假设市场是竞争的。但如果签订契约有沉没成本，情况就并非如此。

我们现在关心这样的市场经济结果：在市场上，个人无法和生产者签订限制其继承人的契约。有许多潜在的生产者。我们假设，签订契约的成本是 ϵ（>0）。根据假设，这是沉没成本。现在很容易检验，是否存在一个子博弈完美均衡：每一个企业索取保留价格 u，并且所有的个人都签订一个 T 年的契约。假设所有的生产者——现在的和未来的——都索取 u，那么一个遵守 T 年契约的代表性个人就不会损失什么，并且会获取 ϵ（$T-1$）——而如果他在一生中的每一个日期都再次进入市场，就必须付出这个成本。因此，对他来说，签订这样的契约就是最优的。假设所有当前存在的个人都签订了这样的契约，每一个日期市场上的个人都是新生的个人。如果企业间的竞争是伯川德竞争，每一个日期不会有超过一个的企业进入市场。因此，在每一个日期都有单个的企业进入，与最新一代的所有人签订一个 T 年的契约，并且每一产品索价 u。企业无法和老一代人签约，因为它们已经和上一个供

41　正如马瑞斯·施瓦兹（Marius Schwartz）在下文中所强调的，"The Nature and Scope of Contestability Theory," *Oxford Economic Papers*, vol. 38, supplement (November 1986), pp. 37－57。还有其他原因导致契约使得市场有效竞争的可能性更小了，而不是更大了。契约使得现有企业承诺，一旦有进入，它就会激烈地竞争；它使得进入是无利可图的这一说法更为可信了。

应商签订了紧的契约。[42]尽管市场上有大量的企业，均衡价格却是垄断价格。

5. 增加潜在竞争者的数量

在上一部分，关于沉没成本和潜在竞争的大部分模型中，我假设了只有一个潜在竞争者。事实上，在伯川德竞争中只需要两个竞争者就可以确保价格和生产的边际成本相等。因此，大部分结果不会因为出现了更多的竞争者而改变。

然而，我已经注意到，在某些情况下，潜在竞争者越多，竞争越不强。原因很简单：如果企业相信，自身的成功进入会鼓动其他企业进入市场，并窃取利润，那么它冒风险进入市场，从现有企业手中窃取部分利润的动机就会减少。事实上，企业可能会认为，如果从现在的现有企业手中抢走客户是值得的，后来的企业也会觉得抢走顾客是有利可图的。我将用两个例子说明这一点。

反应滞后。第一个例子是根据前文讨论的反应滞后的例子稍作调整获得的。在有很多潜在进入者的时候，每一个进入者都会担心，后来的进入者是否会从它手上窃取市场。比如，假设有两个潜在进入者。每一个企业都知道，如果它首先进入，对手会迅速跟随，那么它就不能偿付其沉没成本。因此，它就会推迟进入的时间，直到一个非常接近现有企业可以调整价格的时期，那样其竞争对手就会发现进入是无利可图的。[43]然而，如果有三个潜在竞争者，其中一个进入者会立刻进入，因为它知道，一旦只剩下两个企业，只有在非常接近现有企业可以自由调整价格的事后，才会再有进入。这一论断可以一般化：如果有偶数个潜在进入者，那么直到接近现有企业可以自由改变价格的时候，才会有进入。如果有奇数个潜在进入者，有一个企业会立刻进入。在任何一种情况下，均衡价格都停留在垄断价格水平上。[44]

---

42 我没有讨论违约的罚金问题。均衡契约会使得罚金条款至少足够强到可以在每一个日期阻止违约。因为个人知道，在均衡中，他不可能违约，因此他会毫不犹豫地签订契约。

43 也就是说，如果 L 是任何企业可以改变价格的时间间隔，λ 是进入者进入的时间。则进入者会在 $L-\lambda$ 期间主导市场。通过选择进入时间来得到零利润，也就是说，$\epsilon = \pi(q^m)(1-e^{-r(L-\lambda)})/r(1-e^{-rL})$。

44 当然，这不是唯一的完美均衡。比如，我们考虑下列情况。每一个企业都声称，如果已有奇数个企业进入了，它就不会进入，但是如果已有偶数个（或者零个）企业进入了，则它就会进入。此时，有一个企业进入。企业知道一旦它进入，就没有企业会再进入，因为现在有奇数个企业进入了。这一策略是完美的。假如所有的企业都宣布这一策略，如果在奇数个企业已经进入的情况下，有一个企业进入了，立刻会有企业跟进，因此就会遭受损失（反过来，如果有偶数个企业，企业进入就有好处，因为它知道，它一旦进入，就不会再有企业进入）。还有一个均衡是没有进入；均衡策略和上面描述的策略相同，除了奇数和偶数对调了。

同样，还存在着混合策略均衡，在下文中将有讨论。

如果有有限数量的潜在进入者，就存在一个混合策略均衡：在每单位时间内会有企业进入的概率记作 $\Omega(N)$，它是市场上现有企业的数量（以及它们进入的时间——这指出了现有企业可以再次自由改变价格的时间）的函数。时间单位 Ω 满足，在进入的时候，企业的预期回报正好足以偿付其沉没成本。考虑 L^* 是无穷的这一受限的情况。这样 Ω 就是一个常数，满足 $\epsilon = \pi_m/(r+\Omega)$。

不完美信息和进入阻止。尽管到目前为止，模型假设进入者完全知道它们进入后的情况，然而事实上，信息是不完美的。它们只能猜测现有企业的成本是多少。但是，现有企业对第一个试图与之竞争的企业的反应，揭示了大量的信息，这对于其他的潜在进入者很有意义。如果只有一个潜在进入者，进入是有利的。而如果有很多潜在进入者，就不会有进入了。

假设企业生产的边际成本是高还是低是不清楚的，[45]但是在进入之后，就会爆发伯川德竞争。为了简化，假设潜在进入者的成本函数相同；沉没成本为 ϵ，边际成本为常数 c_e（在最高容量为 Q_e 之前）。因此，如果现有企业的边际成本低于 c_e，进入者就会损失沉没成本 ϵ。如果只有一个潜在进入者，而现有企业的边际成本 $c_h > c_e$，价格等于 c_h，进入者获利 $(c_h - c_e)Q_e/r - \epsilon$，其中 r 表示利率。[46]因此，假如现有企业边际成本很高的概率足够大，就会有进入。我们注意到，在这个模型中，进入会披露信息。如果有无限多的潜在进入者，一旦现有企业的边际成本为 c_h 这一信息变得明显，就会发生进入，直到价格降到和平均成本相等。因此，第一个进入者的总回报仅仅是 $(c_h - c_e)Q_e - \epsilon$。后来的潜在进入者的存在阻止了进入，潜在竞争对于约束市场不那么有效。[47]

三、沉没成本和创新

如果正如我所指出的，沉没成本是普遍的。另外，和很多时候一样，事后竞争（如果不是由伯川德模型来描述的）是相当激烈的。毫无疑问，企业不希望进行价格竞争。企业可能不能理解我构建的多阶段博弈理论模型，

45 Salop，在“Strategic Entry Deterrence”指出，在这样的条件下，现有企业通过它索取的价格来表明自己的边际成本是有好处的。

46 我又回到离散时间模型，其中信息发布存在一期的滞后。

47 当然，这里的问题是，第一个进入者无法独占其行为产生的信息回报。

但是它们知道价格战的结果是不好的。因此，它们试图集中于那些具有潜在利润的竞争形式上：集中于发展新的产品和程序。事实上，正是由于价格竞争的无效率，使得管理层将精力放在研发上。并且提供资源来负担其经费。[48]

在本部分，我们将把沉没成本一节中的结论，扩展到经济部门中创新是很重要的分析中来。我将这一节分为两个部分。首先，我要说明，我先前集中讨论的技术的特征——递增回报和沉没成本——描述了技术变迁。稍后我会证明，在沉没成本上的很小花销，都可以使得现有企业保留它的垄断位置，同时利润不会降为零，并且能对创新的步伐产生不可忽略的影响。潜在竞争可能会几乎完全没有效果。

四、为什么带有技术变迁的市场通常都是不完全竞争的

标准的研究技术变迁的方法，将推动对技术变迁和生产其他产品或服务同等看待。存在一个生产函数，它描述了传统投入要素和产出（比如某种产品）之间的关系，$Q=F(K, L, A)$。其中变量 A 表示知识存量，正如 K 表示物力资本的存量。这样，存在一个生产函数，将知识状态的增加和用于生产知识的投入品——比如，大学教授——联系起来。

$$dA/dt = G(K_R, L_R, A_R, A)$$

其中 A_R 表示关于如何制造知识的知识状态。[49]

我们注意到，传统商品生产的模型和信息生产的模型之间有相似性，因此就存在一个幼稚但是流行的方法，认为知识“市场”像传统商品市场一样运作。这一论断在很多重要的方面都是错误的。首先，技术变迁的生产是非凸的。[50]第二，如果对于固定的技术，是规模报酬不变的（F 函数对于 K 和 L 是规模报酬不变的），那么，如果将对研发的投入考虑进来，

48　回忆我们之前讨论的权益和信用市场的局限性，它通过未分配利润以外的方式限制了企业为研发提供经费的能力。

49　近来，斯蒂格利茨强调了知识状态（关于学习或者学会学习）的重要性，“Learning to Learn, Localized Learning and Technological Process,” in Partha Dasgupta and Paul Stoneman, eds., *Economic Policy and Technological Performance*（Cambridge University Press, 1987）, pp. 125-153。显然，我们需要用一个方程来描述 A_R 的增加。为了简化，在后文中我们都假设 A_R 是固定的，并且在函数 G 中受到了限制。

50　R. Radner and Joseph E. Stiglitz, “A Nonconcavity in the Value of Information,” in Marcel Boyer and Richard Khilstrom, eds., *Bayesian Models in Economic Theory*（Elsevier, 1984）, pp. 33-52。

就是规模报酬递增的了。[51]从企业的角度来看，这意味着信息的价值会随着生产规模递增。一项发明如果能够将制造一件产品的成本降低 $ 1.00，那么如果每年制造 1 000 个产品，那么这项发明就值 $ 1 000，如果每年制造 10 000 个产品，那么这项发明就值 $ 10 000。这一非凸性导致，带有技术变迁的市场天然就是不完全竞争的。在某种意义上，它们是自然垄断的。[52]

我先前讨论了具有自然垄断的均衡的性质——尤其是潜在竞争产生的约束作用——取决于沉没成本的存在。大部分的研发花销都是（根据它的本质）沉没成本。科学家用于研究的资源是无法重新恢复的。一旦他花费了时间，时间就花掉了。[53]同时，我证明了自然垄断即便只有很小的沉没成本，也同竞争行为很不一样。我还指出，认识到沉没成本对事后竞争性质有影响，这导致了企业的沉没成本或者使得企业的沉没成本增多。我认为沉没成本是内生性的。它使得人们投资于研发的动机更强了，因为这些事前花销都是沉没成本，可以用来阻止进入。本节的稍后部分，我将证明，现有企业通过在研究项目上沉没相对小的成本，可以阻止潜在进入者的竞争，甚至可以阻止他们进入研发竞争。

---

51 即便递增回报幅度很小，也会出现自然垄断者。关于沉没成本和潜在竞争的部分证明了，即便沉没成本很小，这样的自然垄断者也能够索取垄断价格。

52 技术变迁还在很多其他方面与传统商品不同。制造技术变迁本质上就是制造信息。因此，技术变迁和传统商品不同的原因，与（更一般的）信息市场和传统商品市场不同的原因是类似的。这些看法已经由斯蒂格利茨给出了详细的论证，“On the Microeconomics of Technical Progress,” in George M. Katz, ed., *Technology Generation in Latin American Manufacturing Industries* (St. Martins Press, 1987), pp. 56–77；以及 Stiglitz, *Information and Economic Analysis* (Oxford University Press, forthcoming)。

传统理论集中关注具有同类产品的市场，而每一项新知识的价格必然和原先制造的知识的价格不同。另外，知识的生产者无法像椅子的生产商可以展示他生产了什么一样，展示它自己生产了什么。因此，信息是很难出售的。同时，很难完全独占技术变迁的回报。（尽管专利法律试图为这种独占提供便利，但它们却是不完全的）。事实上，知识具有很多公共品的性质——要实现排他是很难的，并且也是不合人意的。

最后，对于物质资本的投资通常是通过借贷，并以该物质资本作为抵押品来完成的，而研发花销在很大程度上不是可抵押投资。给定企业筹集资产的能力限制，其原因由 Bruce Greenwald, Joseph E. Stiglitz and Andrew Weiss 在下文中讨论，“Informational Imperfections in the Capital Market and Macroeconomic Fluctuations,” *American Economic Review*, vol. 74 (May 1984, *Papers and Proceedings*, 1983), pp. 194–205，因此，资本限制（信用配给）对研发投资比对传统投资更为重要。

53 技术变迁也常常作为生产的副产品（干中学）而产生。这些生产决策（根据它们的本质）都是沉没成本。Dasgupta and Stiglitz 讨论过沉没成本对市场均衡的重要作用，“Learning by Doing”。

很久以前，熊彼特就指出了研发竞争的重要性（没有使用任何现代词汇）。他没有明确地衡量竞争有效性的福利标准，但是在他的论文中，我们可以发现三条定理：

——研发竞争导致利润下降为零；

——研发竞争导致一连串企业进入，并且竞争持续发生，而不是在某一刻发生；

——竞争使得企业将研发花销定在一个有效率的水平上，从创新中获得的收益，多于研发产生的暂时垄断导致的静态无效。

在以前的研究中，帕萨·达斯古普塔，约瑟夫·斯蒂格利茨，理查德·吉尔伯特以及大卫·纽伯瑞指出，这三条定理都是错误的。[54]他们证明，进入威胁实际上促进了研发，也许超过了社会有效水平，但是现有企业的利润仍然是正的，并且它有动机保持这样的研发速度，以阻止进入，从而维持垄断。

然而，这些研究忽略了沉没成本的重要作用。达斯古普塔和斯蒂格利茨后来指出，在有沉没成本的时候，现有企业可以只花费很少的钱，就能在事先阻止潜在对手。[55]研发竞争可能是相对无效的：利润可能比之前分析所认为的更大，此研发速度会与一个纯粹的垄断者的研发速度稍有不同。潜在竞争无法对市场产生有效的约束作用。

很小的阻止进入的研发花销可以提供完全的保护

考虑一项专利竞争，竞争中的赢家可以获得该发明的所有使用权。在研发技术上没有任何不确定性。为了获得这项专利，企业必须先于其他企业，完成某个固定数量（N）的实验。我假设企业的研发决策总是必须延续两期，并且能力限制使得它们最多进行 M 次实验，其中 $N>M$。令［N/M］表示小于或者等于 N/M 的最大整数。记作 $n^* \equiv [N/M]$。令 $k^* \equiv N-n^*M \geqslant 0$。

假设最初只有一个企业，它不面临真正的或者潜在的竞争。令利率为正。如果对我们正在讨论的产品的需求是增长的，比如说以某个常数增长。假设有一个最优日期 T^*，在那一天现有企业希望完成 N 个实验。由于利率

---

54 Partha Dasgupta and Joseph E. Stiglitz, "Uncertainty, Industrial Structure, and the Speed of R&D", 以及 Richard J. Gilbert and David M. G. Newbery, "Preemptive Patenting and the Persistence of Monopoly," *American Economic Review*, vol. 72 (June 1982), pp. 514-526。

55 以前的模型（没有沉没成本）并不是真正的关于专利竞争的模型，而是一个关于研发资源分配决策的一劳永逸的模型。请参见 Dasgupta and Stiglitz, "Uncertainty, Industrial Structure, and the Speed of R&D"。

是正的，企业将其研究推迟到越接近 T^* 越好。它会在（t，$t+1$）和（$t+1$，$t+2$）期间进行 $e^*(t)$ 次实验，其中

$$e^*(t)=0 \text{ 对于 } t=1,\cdots,T^*-2n^*-4$$
$$k^* \text{ 对于 } t=T^*-2n^*-2$$
$$M \text{ 对于 } t=T^*-2n^*,T^*-2n^*+2,\cdots,T^*-2 \quad (16)$$

现在假设有两个企业，并且它们交替采取行动：企业1在奇数日期选择其研发政策，企业2在偶数日期选择其研发政策。因此企业1有先动优势。现在如果每次实验（或者至少第一个实验）的成本足够小，面临着潜在竞争的企业1就可以先占市场。[56]在有进入威胁的时候，唯一的子博弈完美均衡是：企业1在 $t=1$ 进行 k 次实验，而不是在 $t=T^*-2n^*-2$，在其他日期，其研究计划 $e^*(t)$ 不变。当然，企业2不作任何研究。

这个均衡是由下列策略获得的。现有企业在（1，3）区间，在 $t=1$ 时进行 k 次实验，并且同时宣布在前一个偶数日期，对手进行了多少次实验，它就会在接下来的奇数日期进行多少次实验。假设企业2在某个偶数日期选择进行 m 次实验。除非企业1在下一个日期进行同样多次的实验，否则企业2就不再落后。那么对于企业2而言，采用同样的策略，采取同现有企业同样的行动是有好处的。如果企业2这么做，最糟的结果就是与企业1分享专利——而根据我的假设，这也是有利的。如果它不这么做，其投资就没有回报。因此，如果企业1不能采取和对手同样的行动，它就损失了大量的利润。所以，现有企业采取和对手一样的行动是合理的。因为企业2了解这一点，所以宣布这一策略是可信的，企业2就不会进入。我们用命题5对此进行总结，

命题5：面临着潜在的研发竞争者，现有企业的最优策略（产生了子博弈完美均衡）是：通过承诺在每一个奇数日期 t 进行 $e^{**}(t)$ 个实验，可以预先阻止潜在进入者，其中

$$e^{**}(t)=k,t=1,$$
$$0 \text{ 如果 } S_1(t)<S_2(t)-k,$$
$$M \text{ 如果 } S_2(t)+k>S_1(t)>S_2(t)-k,$$
$$\min[M,S^*(t)-S_1(t)] \text{ 如果 } S_1(t)>S_2(t)+k \quad (17)$$

其中，$S_i(t)=\sum e_i(\tau)$ 是到此为止进行的实验总数，$S^*(t)=\sum e_i^*(\tau)$ 是

56 此处假设获得专利比与对手分享发明的收益多得多，并且与对手分享发明比失去该发明的收益大。

垄断者到 t 时进行的实验总和（这一方程是我用文字描述的策略的正式表述）。

需要注意，如果（$T^* - 2n^*$）和 M 很大，式（17）描述了一个本质上很小的先占优势：现有企业在研发上稍稍领先，而这足以使对手不能加入竞争。[57]

五、结论

传统的经济理论强调，固定成本以及它们引起的非凸性的重要性。非凸性限制了竞争的程度。然而，存在一个广为人知的信念：即便没有真正的竞争，潜在竞争也可能足以约束一个行业中的企业，并且使得价格下降。这一论断要求规模经济是由未沉没的成本造成的。但是，我们还不能清楚地了解，这一结论对假设有多敏感。我已经证明了，任意小的沉没成本都可以作为进入的绝对壁垒，并且使得潜在竞争完全失去约束机制的效力。

同时，我证明了为何这么多现代工业经济中的竞争都集中于创新行为，而不是像传统微观经济分析的那样，集中于价格竞争——至少在这个方向努力有希望获得利润——以及为何潜在竞争在那些研发很重要的经济部门会尤其无效——研发花销本质上就是沉没成本。

本文的大量分析都假设存在伯川德竞争，这样做有两个原因。首先，在伯川德竞争（而不是古诺竞争）下，在真实参与者数量有限的市场上，价

57　关于可能维持垄断的文献指出，为了阻止进入，现有企业必须像其最具进攻性的潜在对手那样，进行高强度的研究，这么做可以获利。例如，参见 Dasgupta and Stiglitz，“Uncertainty, Industrial Structure, and the Speed of R&D”，以及 Gilbert and Newbery，“Preemptive Patenting and the Persistence of Molopoly”。

与之对应的是，在 Dasgupta and Stiglitz, pp. 11-12 中引入了一个非常小的先占优势，从而指出，如果研发要进行好几个时期，那么研发花销是沉没花销这一事实使得现有企业（假设它有先动优势）具有比我们想像中更大的垄断力量。命题 5 证实了这一猜想。

本部分的分析需要和 Fudenberg 以及其他人的研究对照来看，参见“Preemption, Leapfrogging, and Competition in Patent Races”。在他们的分析中，企业必须同时行动，而不是序贯行动。企业集中精力于混合策略完美均衡：后进的企业或者不做任何研究，或者做很多研究，而领先的企业或者保持垄断步伐，或者稍微加速。因此有一定可能，当领先企业保持垄断步伐的时候，后进企业加速研究，从而超过了其对手，并成为新的领导者。而在我们的模型中，超越是不可能的。哪一个模型更合适，这取决于观察和反应的滞后。如果现有企业可以了解到其对手正在进行的实验，并且迅速做出反应，那么本部分的分析（其中企业宣布它会采取和对手一样的行动，但是在过渡时期会保持稍有变动的垄断步伐）似乎是合理的。Christopher Harris and John Vickers，“Perfect Equilibrium in a Model of a Race,” *Review of Economic Studies*, vol. 52 (April 1985), pp. 193-209，将序贯模型分析扩展到更一般的技术。其结果和此处报告的结果相似。

格等于边际成本。因此，伯川德竞争提供了一个自然的基准：毫不惊讶的是，没有伯川德竞争的市场都是没有效率的。第二，竞争性文献的标准“故事”里，每个企业都会给出比别人低的价格，也就是说，作者们似乎都预先假设了竞争采取伯川德形式。

在本文开始部分，我表明本文部分是关于经济学理论，部分是关于我们用来分析经济的模型。此处，我运用了近乎标准的企业互动博弈的完美均衡解的标准概念，某些结果似乎令人很不安。我认为，标准理论传递了三个不同的信息：对进入而言，最基本的就是进入者关于事后竞争本质的信念；事后竞争取决于某些状态变量；进入者关于事后竞争本质的信念是通过一系列自省或者后推获得的——给定博弈的结构和参与者的理性是公共知识，它们通过推算其他企业的理性做法得到关于事后竞争本质的信念。我发现前两个假说相当敏感；第三个假说很值得质疑。[58]真实世界太过于复杂，因此不能被我们的简单的模型很好地描述出来。市场的潜在参与者可能既不是理性的，同时也不确信其对手是理性的。如果潜在进入者通过现有企业的以前行为来对其反应做出预期，然后再进入，现有企业对其的阻止行为，以及现有企业的反应可能同我们猜想的相当不同。如果潜在进入者认为，至少存在很少的机会，现有企业不会像我们猜想的那样做出残忍的反应，假如沉没成本不是太小，它们都会愿意冒险进入。另外，有很小沉没成本的均衡和没有沉没成本的均衡看起来差别不会太大。[59]这些更现实的模型证明了这里描述的结果，它们同时证明了依赖于类似的强理性假说的结果，从而确保了市场经济的效率。

更一般的情况是，我观察到了五个法则，其有效性超越了我们这里探讨的简单模型。

——事后竞争的增加会降低事前竞争的有效性，反之亦然。因此降低事后竞争的政府政策，比如价格管制，可能会实际上增强社会福利，而旨在增加事后竞争的政策会降低社会福利。

——现有企业和进入者都有由其支配的机制，能够影响事后竞争的本

58 即便第三个假说（其作为很多现代博弈论的基础）的逻辑一致性最近也受到了质疑。P. J. Reny, “Common Knowledge and Extensive Form Games,” *Journal of Economic Perspectives* (forthcoming)；以及 K. Binmore, “Modeling Rational Players,” unpublished paper (London School of Economics, 1985)。关于这一研究的综览，请参见 K. Binmore and A. Brandenburger, “Common Knowledge and Game Theory,” *Journal of Economic Perspectives*。

59 瑞查德·迈克马斯特在普林斯顿大学的博士论文中探讨了这些问题。

质。尤其是，现有企业可能会选择沉没成本。在沉没成本和创新部分所讨论的研发问题中，现有企业会通过较早地（比起它是未受限制的垄断者的时候）花费相对小的资金来阻止进入。

——当企业采取阻止进入的行为时，潜在竞争可能会降低福利。[60]因此，在研发上的早期花销并不会导致创新较早出现，并且消费者福利不会改善。但是这些阻止进入的花销确实降低了现有企业利润的贴现值。类似地，在自然垄断的情况下，进入阻止采取过剩能力的形式，但是对消费者没有好处。企业也有可能在没有效率的情况下沉没一定成本，而这仅仅是为了阻止进入。

——真实的进入可能不会增强福利。相反，进入者可能会通过非合作完美均衡的合作安排来容忍进入。现有企业的垄断利润被分享了（以过度的进入花费这一社会成本作为代价），但是消费者并没有得到改善。

——沉没的交易成本会进一步瓦解市场，因此，比起简单看企业和顾客的数量所显示的竞争程度，有效的竞争程度要低得多。

我已经证明了，这些基本的法则是如何解释一些支持竞争理论的非正式论断的。尤其是如下论断：（当存在沉没的生产成本的时候）一旦将不可避免的签订契约的沉没成本考虑进来，追求契约产生的竞争就不那么令人信服了。类似地，价格滞后可能会降低事后竞争，因此会鼓励进入，并让事前竞争更加有效。但是，反应只要有一点滞后，即便小小的进入成本也能阻止进入。如果反应无穷的滞后，进入会发生，但是价格仍然停留在垄断水平：资源由于过度进入而完全浪费了。

这些模型使得政策分析处于两难境地。它们宣扬，依赖市场可能是不够的。它们也表明，在那些递增回报非常重要的部门，解除管制既不会导致消费者价格降低，也不会使得生产更有效率。[61]它们指出，排外领域或者垂直限制之类的行为看似会限制竞争——例如，通过阻止进入或者促进串谋的方式——它们可能实际上正是这样做的。[62]当然，这些模型并不会以研发有重

60　非常重要的是，不要对这一结果进行任何政策含义的解释。尤其是，我并不是倡议降低潜在竞争（并且，事实上，如何来降低潜在竞争并不是显而易见的）。这里所考虑的，仅仅是确知潜在竞争对价格、利润和福利的影响。

61　要了解关于竞争学说对铁路和水路运输公司合并的适用性的讨论，请参见 William B. Tye, "The Applicability of the Theory of Contestable Markets to Rail/Water Carrier Mergers," *Logistics and Transportation Review*, vol. 21 (March 1985), pp. 57 – 76。

62　然而，这些实践可能也有效率收益，其中，人们必须用降低竞争的成本来抵消这样的收益。同时，必须强调的是，这些实践也有可能有效率成本，而不是效率收益。

要的规模报酬效应为理由，来支持那些要求放松执行反托拉斯法的人，并且潜在竞争会确保：即便出现垄断者或者主导企业，企业也需要致力于快速的研发来维持自己的位置。

但是，这些模型同时也指出，那些原本设计用来增加竞争的政府行为，可能会有其自身的成本。因为对于任何自然垄断而言，都有规模经济。这里有一个很明显的解决方法：使用一个基于福利的反托拉斯政策。但是这样的解决方法是一个幻想：反托拉斯法的执行历史强有力地证明了，法律体系并不能提供这样一个框架——其中可以并且将会产生该解决方案的特性。

简单规则的优势是显而易见的。学说本身代表了一个这样的规则。现在有一些人倡议大不相同的、更加宽容的规则。问题是我们已经看到，沉没成本导致了其中的一项得失权衡，尤其是在无法完全实现规模经济的无效率性和使用垄断力量的无效率性之间的得失权衡。[63]如果我们和熊彼特一样相信，由竞争所提供的创新动力起着中心作用，并且相信那些声称小企业在创新过程中有重要作用的研究，那么严格的反托拉斯法带来的收益，就会大于其带来的损失。但是我相信，这仍然是一个有争议的问题。

附录Ⅰ：生产单个产品的进入——退出博弈的其他构建方式

在另一种进入——退出博弈的构建中，在每一期中有四次行动。第一，进入企业决定是否进入，或者，如果它已经在市场上了，它要决定是否退出。然后，现有企业决定是否退出。接着，进入企业设定价格。最后现有企业设定价格。行动的顺序是人为设定的。其结果和文中描述的同时行动博弈相同，但是稍微简单一些。很清楚，在任何一个有两个企业同时存在的时期，均衡价格是伯川德价格；因为，如果进入者设定的价格高于伯川德价格，它就无法售出产品，因为现有企业出价会比它低。[64]同时清楚的是，在

63 在本文的很多观点中，我已经提醒读者注意阻止进入战略导致的无效率问题。然而，这些行为在严格的反托拉斯政策下更好，还是在很松的反托拉斯政策下更好，这是不明显的。在严格的政策下，一个大企业可能会对从事这些行为非常犹豫，免得常常受到反托拉斯控诉；另外，市场上存在更多的企业可能意味着进入阻止行为不是很有效。在很松的反托拉斯政策下，潜在竞争和真实竞争可能会减少，因此对进入阻止政策的需求也会降低。

64 进入者事实上对于其索取的价格是无差异的，因为它知道，或者是对方出价比自己低，或者是它除去可变成本后获得零利润。但是，如果进入者产生了一定成本，无论它是否售出了第一个单位的产品（也就是说，要维持企业会产生一定成本，即便它不生产），唯一的均衡就是我们在文中描述的那个。

第三期，如果进入者不退出，当且仅当 $F_1 \geqslant 0$ 的时候，现有企业发现退出是有利的（如果 $F_I = 0$，现有企业对是否退出是无差异的）。进入者了解这一点，在第三期当且仅当 $F_I \leqslant 0$，$F_e \geqslant 0$ 时，它才会退出）。

倒推一个时期，如果 $F_I > 0$，若第二期发生进入，现有企业会立刻退出。因此进入会发生，但是，结果是由一个垄断者代替了另一个垄断者。如果 $F_I < 0$，则没有进入。再一次地，进入威胁对价格没有任何影响。

附录Ⅱ：有退出的多期博弈

当期数得到扩展的时候，我们可以得到更多的合作均衡。出于同样原因，现有企业要阻止进入就更难了。

为证明这一点，考虑一个四期模型。第三期合作的条件和前文分析的一样，并且假设这些条件是满足的。令 α_2 表示第二期现有企业分配到的垄断利润份额。只要下式得到满足，现有企业合作就是有好处的，

$$\alpha_2 \pi + \delta^2 g^2 \pi \geqslant \pi + \delta F_I$$

也就是说，只要满足

$$\alpha_2 \geqslant 1 - \delta^2 g^2 + (F_I \delta / \pi)$$

类似地，要让进入者合作，就必须

$$(1 - \alpha_2)\pi + \delta g \pi + \delta^2 F_e \geqslant \pi + \delta F_e$$

也就是说

$$\alpha_2 \leqslant g\delta - \delta(1 - \delta) F_e / \pi$$

很容易验证，这两个条件可能是相容的，并且提供了一系列和完美均衡一致的 α_2 的值。比如，如果 $g = \delta = 1$，合作要求 $1 \geqslant \alpha_2 \geqslant F_I / \pi$。

评论和讨论

丹尼尔·麦克法登（Daniel McFadden）：约瑟夫·斯蒂格利茨的论文的主要思想是：沉没成本是普遍存在的，并且现有企业可以借此来预先阻止进入者。特别是，对于研发的投资可以当作一种沉没成本，使得现有企业可以索取垄断价格，而不用担心其他企业会进入。论文回顾了关于沉没成本和竞争性的主要争论。第一，如果有正的沉没成本，并且即使伯川德竞争是可

信的，那么竞争性就不存在了。[65]第二，在需要通过沉没过剩能力来使得伯川德竞争威胁是可信的时候，均衡可能要求容忍小的进入者，并且限制价格。[66]第三，如果协商契约本身有沉没成本，那么通过事先签约来消除真正的沉没成本是不可行的。我要增加的观点是，使得现有企业的反应变慢以及鼓励进入的契约是没有吸引力的。相反，它们会寻求那些一旦进入发生，就会使它们进入伯川德竞争的契约。

在这篇文章最具创新性的部分，斯蒂格利茨把研发看作一种沉没成本。他认为即便沉没很小的成本来获取微小的技术优势，也可以制造出可信的威胁，来阻止进入。从而就有一个进入专利竞争的模型。[67]斯蒂格利茨的模型的重要特点是：现有企业和进入者是序贯行动的，现有企业首先行动。没有外生的不确定性，每一个参与者对所有已发生的行动都拥有完全信息。在知道要获取将产权转化为生产产品的专利所需的累积投资水平的情况下，参与者的行动就是决定研发投资水平。模型同时假设，无法基于和企业科学家签订的专利合约，通过剥削劳动力或者榨取租金来偿付研发成本；因此，这些成本是沉没的。最后，模型假设研发投资能力是有限的，因此需要两个或者更多的时期来获取一项专利。

斯蒂格利茨在模型中给出结论是现有企业有更优的策略可以阻止进入：现有企业最初有一笔研发投资，此后进入者有多大研发投资，现有企业就也给出多大研发投资，从而确保赢得专利竞争。因此，这里存在一个子博弈完

---

65 请参见 Partha Dasgupta and Joseph E. Stiglitz, "Uncertainty, Industrial Structure, and the Speed of R&D," *Bell Journal of Economics*, vol. 11 (Spring 1980), pp. 1–28；Richard J. Gilbert and David M. G. Newbery, "Preemptive Patenting and the Persistence of Monopoly," *American Economic Review*, vol. 72 (June 1982), pp. 514–526；以及 Joseph Farrell and Garth Saloner, "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation," *American Economic Review*, vol. 76 (December 1986), pp. 940–955。

66 请参见 Steven C. Sallop, "Strategic Entry Deterrence," *American Economic Review*, vol. 69 (May 1979, *Papers and Proceedings*, 1978), pp. 335–338；Avinash Dixit, "The Role of Investment in Entry Deterrence," *Economic Journal*, vol. 90 (March 1980), pp. 95–106；以及 Richard J. Gilbert, "Preemptive Competition," in Joseph E. Stiglitz and G. Frank Matthews, eds., on *New Developments in the Analysis of Market Structure* (MIT Press, 1986), pp. 90–123。

67 达斯古普塔和斯蒂格利茨曾经建立过相关的模型，请参见：Dasgupta and Stiglitz, "Uncertainty, Industrial Structure and the Speed of R&D", Gilbert and Newbery, "Preemptive Patenting and the Persistence of Monopoly"；以及 Drew Fudenberg and others, "Preemption, Leapfrogging, and Competition in Patent Races," *European Economic Review*, vol. 22 (June 1983), pp. 3–31。

美均衡：现有企业通过很小的沉没的研发投资来预先阻止进入。因为进入没有发生，获取专利所必需的研发投资不会彻底完成。

很自然的问题是，斯蒂格利茨的结论是否依赖于该模型的特别假设。首先，考虑同时行动的可能性（或者等价的，采取序贯行动，但是第二个参与者不知道第一个参与者已采取的行动）。在这种情况下，参与者会采取随机策略，但是通过研发投资可以预先阻止进入，这一点仍然成立。[68]

第二，考虑研发技术。研发投资得到的专利可能会降低生产的边际成本，增加产品质量或者引入产品差异化，因此增加了产品需求，或者将专利产权转化为产品生产。斯蒂格利茨考虑了最后一种情况，其中研发没有直接的社会回报。因此，不仅早期先发制人的研发投资有社会成本，而且当先发制人的研发投资阻止了进入之后，现有企业采取的垄断价格也是有社会成本的。如果产品的边际成本降低了，或者产品质量提高了，那么研发就有一个适当的并且净的、直接的社会回报，并且现有企业的行为以及社会结果都需要进一步地检验。

如果研发行为涉及建立成本，那么以稳定的速率进行长期投资，比用相同的累积投资量收放循环会更有生产力。这样的循环使得现有企业的先占行为花销更大了。如果专利有一个私有汇报，就会打乱平衡，促使研发早日完成。然而，研发实验的规模不经济和范围不经济，可能会使得进入者不容易赶超。[69]

实验成果的外生不确定性是大部分研发行为的特征。不确定性是如何影响研发先占行为的可能性的？从直觉上来看，先占行为的可能性是大大受限的。这样使研发领先可确保现有企业获胜的领先地位（knife－edge）消失了。现有企业先研发的可能效应是创造了一个领先地位，然而，进入者可能会发现进入专利竞争的期望利润比起它在其他行业的机会要好。这个情况可能要求现有企业在研发先占上的投入足够大，使得进入者的期望收益降低到小于其他机会的收益的程度。这种情况下的均衡也相当地依赖于参与者获得的信息。如果潜在进入者知道了现有企业实验的成功程度，进入者就会在现有企业处于不利地位的时候进入。这反过来也迫使在位者采取行动，增加事后先占的结果的可能性。

最后一个关于研发技术的问题，是关于在多产品市场上专利和先占的地

68 Fudenberg and others，"Preemption，Leapfrogging，and Competition in Patent Races"．

69 如上考虑了一些围绕长期研发契约的生产力情况。

位问题。我们可以考虑这样一种技术，它的形状好像一棵树，基础生产过程好像树干，支撑许多种产品的生产，专业化生产过程好像树枝，仅仅在末端影响少数产品。专利加强了这些生产过程，并且允许专利拥有者，通过垄断或者特许他人生产该专利涉及的产品，来独自享有回报。控制技术并且某些分支上面临进入者的现有企业会如何开展研发？先占会将研发力量转移到受到威胁的分支上去吗？那些没有受到进入威胁的产品的质量和成本受到了怎样的影响？

最后一组围绕竞争性的问题考虑这样一种情况：研发是对社会有益的，比如说可以降低边际成本。当且仅当研发受到社会欢迎的时候（如果必要，假设"一揽子"重新分配），一个可以采取完全歧视政策，并且独占所有的研发收益的垄断者才会作此投资。假设该企业现在面临着潜在进入。该行业的研发水平是否仍然会接近社会最优水平？下面的推理证明答案是否定的。如果不能有效地利用研发来预先阻止进入以及由此引发的竞争，现有企业就难以确保自己独占研发收益，因此其开展研发的动力就下降了。这在什么情况下，会产生取决于其对手可以获得的信息以及对其对手威胁的可信性，并且它不大可能和研发的社会有益水平相吻合。另外，现有企业预先阻止进入的好处，可能会导致研发对社会无益，如斯蒂格利茨的模型中展示的那样。然而，如果研发是对社会有益的，并且有限的研发可以导致先占，但是潜在进入者的存在仍然会降低在位者独占利润的能力，这一情况可能导致研发水平低于其社会最优水平。这表明促进竞争市场结构的政策，对于区分对社会有益的研发和对社会无益的研发是非常没有效率的。

山姆·派尔兹曼（Sam Peltzman）：我发现可以用两种眼光来看这篇论文。第一种是对于竞争性结果逻辑上所要求的最低条件的一个调查。第二种是对于在实践中可能导致价格和边际成本分离的情况的详尽描述。作者并没有告诉我应当采取哪种观点来看待这篇文章，但是这篇文章作为一种意在讨论某些本质性问题的逻辑练习来说，是令人印象深刻的。作为一种逻辑练习，它没有什么问题，尽管我认为其结论太强，并且自相矛盾。然而，我对其本质信息有更多的保留意见，尽管这些保留意见是具有方法论意义的，并不是针对这一点的。

作者的论断依赖于标准理论里关于轻易进入所引起的领先地位。如果期望利润是非常小的正数，每个人都希望进入市场。如果该期望利润是非常小的负数，那么潜在进入者就会不进入。现在假设进入有微小的沉没成本，将其加入到完全竞争的情况中，并且使得进入者相信竞争会相当激烈，使得价

格在进入后立刻降到和边际成本相等。极小的利润此时变成负的，因此进入被阻止了。非常重要的是，完全垄断价格可以得到维持。这一度被称为毁灭性竞争。毁灭性竞争理论认为损失很小，但却是毁灭性的。此处，对毁灭性竞争（价格<事前平均成本）的预期完全阻止了进入，并且支持了垄断价格。结论是最简单的可持续结果（价格等平均成本）依赖于这一领先地位。价格并不像之前人们相信的那样，会随着沉没成本的增加而渐渐减少。

斯蒂格利茨从中得到了一个更一般的结论——一个早已存在的结论——对于激烈的、即时的、价格等于边际成本的竞争的预期会降低真实的竞争程度，并且更一般地，阻止了可以增进社会福利的投资。对于可持续性的特别结果和价格等于边际成本的竞争的潜在问题的宽泛含义，我没有发现可以指责的地方。比如，后者就是对前文所提到的知识的排他使用的不良性，这一陈腐观点的对抗，并且成为其他可能的情况的基础——这些可能的情况描述了，在短期内减小竞争来获得帕累托较优资源配置所获得的潜在福利收益。例证还包括，旧的幼稚工业需要保护的论调，以及现代的可持续性要求进入限制的论调。这些情况都引发了我们的疑问，这些限制在实践中怎样发挥作用？但是没有一种情况，表面上看来是不真实的。

作者关于可持续性的特别结果是一种有用的提示，因为进入——退出是为数不多的领先地位结果起重要作用的理论领域之一，我们不应当只就连续性结果达成一致。

如果我们建议斯蒂格利茨在普林斯顿的同事就此问题给出回应，他们会说："嗯，我们对此类问题通常观点一致；这个问题有什么特殊性？"例如，我的学生有时会给出这样的评论："如果利润是很小的正数，每个人都希望进入这个行业。于是每个人都进入了，但是如果所有人都进入了，利润就是负的了。所以，如果它们事前想到这一点，就不会有人进入。您对此有何评价？"我的回答是："除了在一些不便的情况下，一般来说，经济学是一种理想状态。"它们不会都进入。这就是我们可以得到很好的性质或者竞争的原因。

在文章的结尾部分，斯蒂格利茨试图预期回应，并且试图（本不必要的）过度扩张自己的结论。首先，他预期到一种反驳：垄断价格均衡可以被长期契约破坏。他采用一个特殊的模型来笨拙地弥补这一点，在该模型中只提供垄断价格的契约。对我而言，这只会引发理论上的不经济的竞争。我可以立即引用科斯理论。如果一个模型产生了未被利用的收益，其他的模型会随之通过某些交易来实现这些收益。在文中的情况下，出人意料地存在着

永远地面临垄断价格的买家，而潜在卖家愿意以更低的价格出售，并且在此情况下仍然获利。因此，我们可以预期到一种反驳的论调：存在一种交易可以从大体上打破垄断价格，并且使买家和潜在卖家同时获利。

我无法预测这一交易会是怎样的。也许可以一部分买家联合，承诺支付沉没成本，增加一点垂直整合，达成协议，这也许会获得成功。我认为斯蒂格利茨也不能预先给出回应。我认为合适的回应可能是这样的。如果打破垄断价格需要些小伎俩，比如长期契约，那么可持续性就成为一种特殊情况，其中无论合适的交易会是怎样的，其所需的交易成本都是微不足道的。

关于他讨论的反应滞后问题，我的回应与他的讨论类似。反应滞后并不像长期契约那样，它们是斯蒂格利茨可以完全忽略的外在的考虑因素。它们终究促成了伯川德对古诺的批评。斯蒂格利茨在结尾部分非常不自然地处理了这一问题：在某些情况下，它们有助于限制垄断力量，但在有些情况下，它们不起这个作用；另一方面，它们可能根本就不重要。

这一方法有两大难题。一个是理论上的。这里的反应滞后是武断地强加的。斯蒂格利茨无法决定究竟应当采取多么武断的方式来强加反应滞后。给定这种方法，它们的角色仍然是模糊的。最终，对反应速度的限制一定来源于不确定性以及侦查和估计对手的行动后果的成本。它们不会是任意的。更重要的是，简单地忽略反应滞后的重要性，损害了本文所传达的更重大的信息（该信息是我所认同的）。反应滞后是对进入后竞争的激烈性的一种限制，如果斯蒂格利茨简单地提及：“这正是我的观点。这样的摩擦常常是有效竞争所必需的”，就不会损失任何信息了。

现在，我要转而讨论这篇文章的观点对于真实市场应当如何运作的指导性。此处，我必须运用一些想像力，因为本文只是暗示性地提到了文章对真实行为的应用。一个可能的推论是，进入阻止，事实上很重要，它是少数成员竞争的普遍特征——因为沉没成本是如此普遍。然而，我不知道有任何令人信服的证据表明这一点是对的，而我对数据的粗略浏览也并没有证实这一观点。我所想到的一种构建方法是：是否进入越少，市场离垄断就越近？这样的修正是必需的，因为这一理论所描述的垄断在真实世界中并不存在，除非在受管制的市场上。无论如何，我将斯蒂格利茨的理论中描述的垄断，定义为少数成员的竞争，从而推翻了标准的行业分类。我定义了两个最高集中度的行业和两个最低集中度的行业——其标准是行业内最大四个企业总份额高于60%或者低于20%——四位数企业在两位数企业中的比例。然后，我计算了进入率。以下是计算结果：

	高（$CR>60$）	低（$CR<20$）
$100\times\overline{\Delta\log N}$	4.6（24.8）	2.0（32.2）
$100\times\overline{\Delta\log VS}$	71.6（45.0）	76.5（28.8）

第一行表示从1972年到1982年，企业数量的平均对数变化。括号里的数值是标准差。第二行表示同样行业的货物价值的平均增长率。高集中度的行业和低集中度的行业的增长速度几乎同样快。因此在两种行业里，新的进入者的空间差不多大，平均进入率也是相同的。另外，如标准差所示，高进入率的几率也是几乎相同的。现在，任何看似粗糙的东西都是可以确定的；我所要说的是，在真实世界里，需要进行战略性的进入阻止。

斯蒂格利茨认为需要真实地进入来降低价格，而不是仅仅潜在竞争就可以降低价格，并且引用了航空业的例子。我阅读了目前关于航空业管制经验的文献，文献表明斯蒂格利茨的观点是正确的。确实需要新的进入者才可以显著地降低机票价格。过去的关于利润集中的主要文献，提供了一个更间接但也更全面的测试，它们也趋向于支持斯蒂格利茨的观点。但是斯蒂格利茨在普林斯顿的同事也可以说，这些文献支持了竞争理论。

例如，我们假设一个高度集中的行业——一个有着80%集中度的行业——遭遇了大量新进入者，它们成功地获取了大企业一半的业务。合理的典型利润集中分析表明，给定这些进入者，价格将下降5%。这样的结论对于那些支持竞争性的人来说不是太坏，即便它是在向这样一个方向推进——斯蒂格利茨似乎希望人们相信世界是向这个方向演进的。

或者，它可以认为成员很少的行业（接近垄断市场）运作得更好，因为它们可以期望避免激烈的竞争。再一次地，关于产业组织的实证文献倾向于支持这一观点：高度的、上升的集中度似乎和动态更优的价格、生产业绩相挂钩。但是集中度的显著减少，倾向于和更优的价格、生产业绩相连。所以，数据可能仅仅意味着，技术的大幅进步会颠覆已有的结构。

因此实证文献的大意是，市场结构确实起作用，其发挥作用的方式大致正如这篇文章所示，但是它并不像关注这一问题的理论所认为的那么有用。

然而，可能这些和这篇文章都没有密切的关系，它可以支持我刚才得到的几乎所有推论的否命题。真实进入会降低价格吗？有一个模型认为进入导致了串谋。另一个模型认为它导致了浪费的重复投资。成功地阻止进入是否会促进效率？一系列模型都认为，它导致了相反的结果——过剩能力，过度快速地退化等。斯蒂格利茨可能会回答——在某种意义上是合适的——每一

种结果都是有可能的，尽管是在不同的情况下。我对此并不否认。我反对的理由是：我们并没有对在什么情况下能产生这种结果，或者什么情况下会产生那种结果给出任何指导。

因此，我对这篇文章的意见并不像对其主张的意见那么多。早先时候，在过去十年对产业组织文献的主要见解是：其强调对竞争性质有作用的因素是，一些影响企业如何战略性地互动的变量。毫无争议，非常清楚的是：在过去十年间，众多建模者将精力投入到对战略互动的研究中。但是尚未证明的是，在探讨它们的模型之外，吸引人们投入巨大精力的战略考虑是如何（甚至是不）起作用的？

综合讨论

参会者主要关注的是，本文的众多模型在指导实践工作中所起的作用。罗伯特·维利戈（Robert Willig）提出，这篇文章的意义在于：现有企业有时候有措施可以阻止进入，有时候没有。他认为，这篇文章应当被视为一种指导——指导人们找出，在什么情况下，哪条路走不通。

卡尔·夏皮罗（Carl Shapiro）指出，这篇文章引发我们关注两个对预测市场如何运作尤其重要的因素。一个是沉没成本的程度，这决定了进入市场有多难或者退出市场有多难；另一个是当市场上有多于一个企业时，可以预期到的行为。采取即时削价（伯川德）竞争这一极端假设，即便很小的沉没成本也是至关重要的。另一个极端——没有沉没成本，且没有在位者的价格回应——得到了完全不同的结论。他提出，建模的一个作用就是研究结果在接近极端时候的灵敏性。

理查德·莱温（Richard Levin）注意到，对两类变量的效应的区别已经成为某些关于竞争性的实证研究的基础。比如，马格丽特·皮特瑞夫（Margaret Peteraf）近期关于航空市场垄断的研究，试图衡量价格被两类变量降到了什么样的程度，一类衡量沉没成本的程度，另一类试图依据过去的声誉来预测削价行为。类似地，克里夫特·温斯顿（Clifford Winston）注意到，关于航空业的实证研究发现，潜在竞争似乎有作用，但没有真实竞争的作用大。

约瑟夫·斯蒂格利茨的回应是，本文并不是意图提供具体的实证建议，而是提供一种框架来对市场分类——依据某种退出成本以及预期这些成本将会带来的事后行为。他指出，这些模型故意过分简化了，每个模型都只有三到四个参数，但是它们产生了非常不同的行为。所以，它们提供了一种框

架，帮助我们区分在什么条件下会引出什么样的回应。

蒂莫西·布莱斯纳汉（Timothy Bresnahan）观察到，关于这篇文章是否提供了经得起检验的实证建议的争论，更深入地来看，其实是关于理论在实证研究中的作用的争论。他认为，我们不一定去要求理论家给出实证假说。如果一定要的话，可能就得要求他们失去想像力，以至于不能创造出很好的理论，同时，这一要求鼓励实证研究者们在阅读理论的时候也没有想象力。山姆·派尔兹曼同意这一点，理论家们并不一定要创造出精确的实证预测。但是他的批评更加广泛，他认为，到目前为止，关于战略互动的所有研究都缺乏新的洞察力，尽管所有的资源都投入到企业中去了。

考虑到这一点，温斯顿建议大会所有参与者总结关于竞争性的争论的现状。维利戈认为，争论致力于将理论家看作一个深思熟虑的潜在进入者，强迫他们考虑情形、风险以及对现有企业的反应的合理预期的结构性事实，然后询问怎样的价格——成本离差才能让特定的进入计划有利可图。他认为，类似这篇文章中介绍的那些模型，能够使得分析对于不同的条件和不同的行为非常敏感。

斯蒂格利茨认同这一点：博弈论模型方法本质上是内省的，因为它求证状态变量是怎样的，以及给定面对这样的状态变量，理性人会如何行动。有鉴于此，他承认：关于人们依赖于幼稚的模型，来预期其对手的行为（比如通过观察他们是否会报复这一声誉）的证据，对于近期的理论研究是有扰乱性的。

迄今为止，夏皮罗观察到，争论主要是关于进入前价格和进入后的定价行为是否有关的问题。如果二者存在很强的联系，竞争性理论就是适用的。如果有很弱的联系，潜在进入者就不能影响到当前价格。

干中学、市场结构与产业政策和贸易政策*

“只有通过实际行动才能学会那些应该要学会的事情：人们通过砌房子来成为建筑工匠，通过演奏乐器来成为乐师。”

亚里士多德，*The Nicomachean Ethics*，J. A. K. Thomson 译，Penguin Books（London），1976，p. 63。

一、引言

在许多产业中，从经验中学习是非常重要的，尤其是在企业发展的初期。生产率的提高，不但是在资本积累和研发活动上配置资源的结果，而且还是生产过程的一种副产品。也就是说，干中学也能提高生产率。[1]

学习使得生产中产生了一种特别的跨期外部性（intertemporal externality）。通常会将这种现象视为保护幼稚产业的论据。这种观点认为，若政府不进行干预，来自外国的竞争会抑制本国具有学习能力的幼稚产业的发展（对这种分析及其局限性的精彩介绍，参见 Baldwin，1969；Wan and Clemhout，1970；Negishi，1972）。

就我们所知，阐述“幼稚产业论”的文献总是假设，外国的竞争者组成了一个竞争性的产业，而且在这个产业中学习已经停止了。于是，关注的

---

* “Learning-by-Doing，Market Structure and Industrial and Trade Policies”，with Partha Dasgupta，Oxford Economic Papers，New Series，Vol. 40，No. 2（Jun.，1988），pp. 246－268. 本文的研究始于1984年春季。当时，达斯古普塔为斯坦福大学经济学系的访问教授，斯蒂格利茨为斯坦福大学胡佛研究所的访问教授。在1987年春季完成本文的研究时，达斯古普塔为哈佛大学的访问教授和国家经济研究局的访问学者。衷心感谢这些机构让我们的访问得以成行，并使访问非常愉悦。我们与保罗·戴维（Paul David）就本文的主题进行了多次交流，从中受益良多，在此谨致谢忱。非常感谢吉恩·格罗斯曼（Gene Grossman），提姆·克奥（Tim Kehoe），卡拉·克瑞施纳（Kala Krishna），埃里克·马斯金（Eric Maskin），哈米德·萨鲍润（Hamid Sabourian），迈克·温斯顿（Mike Whinston），以及匿名审稿人对本文初稿的深刻评论。英国经济和社会研究委员会（the UK Economic and Social Research Council）与美国国家科学基金会（the US National Science Foundation）提供的资助，使我们可以持续地研究技术变迁与市场结构这一问题。

1 对学习曲线的经验研究包括，阿舍尔（Asher，1956）和阿尔钦（Alchain，1963）对机身生产的富有启发性的分析，齐默曼（Zimmerman，1983）对核电技术的研究，以及利伯曼（Lieberman，1984）对化学加工工业中的生产和投资的分析。荷兰德（Hollander，1965）研究了研发技术中的学习现象。

焦点就集中在本国的学习之上。[2] 但是，这些文献有一个很大的缺陷：它们并没有探讨学习效应会怎样影响本国的产业结构？如果不理解这一问题，那么就无法讨论政府的政策是否合适。本文的主要目的就是研究这一问题，并且以我们的研究结论为基础，分析几种广为人知的产业政策和贸易政策。

干中学与一种沉没成本有关。能增加经验的生产活动就是沉没成本。因此，学习本身就表明了生产活动的不可逆性。在本文的分析中，有一个相关的特征非常关键：除非在所有相互竞争的生产单位之间，在瞬间学习就能无成本地完全溢出，或者除非每期各生产单位在一定的产出水平上，规模收益都严重递减，否则超出一个生产单位，就会造成社会资源的浪费。简而言之，该产业是自然垄断产业。为了把这一问题阐述清楚，假设某一个产业由社会统一管理。在该产业中，各生产单位有学习的潜力。同时假设学习不能完全地溢出。而且，在任何一个生产期间，各生产单位的平均生产成本都独立于该期的生产规模。在这些条件下，从整个社会的角度看，宁愿让一个生产单位在一定时期内生产特定数量的产品，而不愿让两个生产单位去生产。若由两个单位去生产，则每个单位在每一期都只能生产该产业一半的产品。这样，在后一种情形中，每个单位就学得更少了。这是生产中的动态规模经济。它表明，如果该产业处于私人部门中，那么就会形成寡头垄断结构，而不是竞争性市场结构。至少在产业发展的初始阶段（此时学习的空间很大）是这样的。

这些观点都是非常简单的。可能不那么简单明了的是，企业特定的学习提高产业集中度的可能性。具体而言，我们可以预知，很强的学习潜力，再加上激烈的竞争，就会使历史变得很重要（Atkinson and Stiglitz，1969）。此处讲的历史指的是，如果一家企业开始时就获得了一些优势，那么它就可以通过吞食竞争对手的市场来将这些优势资本化。它的优势会随着时间的推移而增长，并使得竞争对手在长期无法有效地参与竞争，除非竞争对手能获得技术创新，或者是除非优势企业的学习能力下降。[3] 从另外一个角度看，我们

2　我们只关注那些以干中学为分析基础的幼稚产业文献，而不去考虑其他文献。这些文献的分析基础是，与学习无关的企业间外部性，以及劳动和资本市场的扭曲。对这些观点的综述，参见 Haberler（1987）。

3　关于在更一般的背景下对这一观点的数学探讨，参见如下著作：Arthur（1983）以及 Arthur, Eromoliev and Kaniovski（1986）。不久前，冈纳·缪尔达尔（Gunnar Myrdal），保罗·戴维以及尼古拉斯·卡尔多（Nicholas Kaldor）在研究发生在某些路径依赖过程中的累积因果关系时，就强调了这一观点。例如，参见 Myrdal（1957），David，（1985，1987）和 Kaldor（1985）。当然，在演化生物学中，也有累积因果关系的观点。对此的精彩解释，参见 Dawkins（1986，chapter8）。

可以推测，最终产品的生产本身就是一种先占行动，可以用它来阻止其他企业进入该产业，或者在其他一些条件下，使留在该产业中的竞争对手的利润越来越少。再从另一角度看，很有可能会将“学习”用于创造进入壁垒(Scherer，1980，pp. 250-252)，或者在其他一些条件下，将学习用于阻止竞争对手留在产业中。到底会出现哪种情况，这取决于学习的空间是否很大，以及各企业是否具有远见。不过，这些问题的答案都并不明显。必须要注意到，即使在初始阶段学习的空间很大，最终也可以将其忽略不计：毕竟，单位生产成本不可能小于零。因此，如果寡头垄断企业富有远见，那么处于优势地位企业的先占空间就会变小。其原因在于，处于劣势地位的企业知道，如果它们耐心一点，并生产足够多的产品，那么在将来就会获得寡头垄断利润。但是，如果各企业以一个很高的贴现率来贴现未来的利润，那么这一观点就不会成立，而且产业集中度会不断提高。

正是由于存在这些前提条件，所以就可以推知，斯彭斯（Spence，1981）以及弗登博格和梯若尔（Fudenberg and Tirole，1983）对存在干中学的寡头垄断产业的分析，并不完全正确。首先，他们都假定产业中存在一定数量的企业（因此，产业结构就是给定不变的）。其次，在他们所研究的产业中，所有企业都是一样的。再次，他们只分析了对称的市场均衡。也就是说，在他们所分析的均衡状态中，所有寡头垄断企业的行为都是相同的。前面我们已经指出，这些假定很可能都是一些不常见的特征。下面，我们将会以多种方式去阐述一个产业，以表明很强的学习潜力确实会导致产业集中的形成。因此，本文最初的实证分析（第四部分）的主要目的就是要表明，当存在很强的学习潜力时，对称的结果就是令人误解的概念。为了集中关注存在学习时出现的动态规模经济，我们假设在每一期各企业的规模报酬都不变。这一假设与斯彭斯以及弗登博格和梯若尔的文章所做的假设相同。我们将会在第四部分的第 2 点研究产业集中的形成。与斯彭斯以及弗登博格和梯若尔一样，在第四部分的第 2 点的分析中所用到的例子，也假设数量给定的几家企业在数量竞争中进行古诺博弈。然而，与前面几位作者不同，我们将会假设，有一家企业最初就拥有成本优势。这也许是一个很小的成本优势。我们的分析将会表明，如果学习的空间很大，而且企业以一个很高的贴现率去贴现未来的利润，那么最初的优势就会随着时间的推移而增加。这体现为优势企业的市场份额会不断增加。在这些条件下，产业集中度就会不断提高，最终会形成垄断（命题 3）。从分析产业演化的角度看，在这些条件下，假定所有企业都是相同

的，那就是一个不好的假设。

当只有少数几家企业展开数量竞争时，竞争不会很激烈。因此，在第五部分，我们会研究一些面临学习曲线的企业激烈竞争的结果。要将这一点进行模型化，最便捷的做法就是，让所有企业就同质的商品展开价格竞争，而不是进行数量竞争。在第五部分，我们将会采用这种类型的竞争。实际上，在第五部分我们还会关注另一个问题——进入。我们以一个简单的例子展开在第五部分的分析。在该例子中，产业中企业的数目是给定的。假设开始时各企业都面临相同的（单位）生产成本。进一步假设，其中一家企业会以消费者忠诚的方式获得一个优势。这将会是不对称的源泉。我们的分析将会表明，在动态均衡中，受到零利润条件约束的优势企业，将会是仅有的一家进行生产的企业。虽然此处的分析与可竞争市场的思想（Baumol, Panzer and Willig, 1982）类似，但是在解释这些结论时，则存在重大差异。由于我们的模型考察的是一个由学习引起的显性的跨时结构，因此垄断企业的利润流就不是一个常数。考虑到未来的利润，为了阻止竞争对手进行生产，垄断企业在开始一段时间会遭受一些损失。在动态均衡中，利润通常会随着时间的推移而增加，最终可能会上升至垄断水平。

这一特征值得重视。当存在学习时，垄断企业无法获得保持不变的利润流。产业并非处于静止状态。垄断企业在成熟阶段能获得正利润的事实，并不能成为假设它一直都能获得正利润的理由，特别是不能成为假定它过去也不曾遭受损失的理由。正如第五部分第1点中的分析所表明的那样，在我们的模型中，优势企业为了取得垄断地位并获得垄断利润，它会在开始时将产品价格定在（当时的）平均生产成本之下。

第五部分第1点中的模型有一个缺点，就是预先假定所有企业都处于产业之中。因此，在第五部分第2点中我们假设，最初产业中只有一家优势企业进行生产。我们将会问：竞争对手的进入威胁能否限制现有企业的市场势力。在命题5中，我们会注意到，如果其他企业可以自由进入，那么最终的结果将无法预料。也就是说，将会出现多重均衡。在多重均衡中，一极是现有企业像一个不受约束的完全垄断企业那样行事，另一极则是垄断企业获得的以现值衡量的利润会减少至零。然后，（在命题6中）我们将会指出，这是一种似是而非的多重性：引入正的进入成本后，哪怕进入成本很少，都会剔除掉其他均衡，而只剩下第一个均衡，即现有企业像一个不受约束的完全垄断企业那样行事。在这里，“看不见的手”不只是软弱无力，而是已经瘫痪

了。该产业成了一个不受约束的完全垄断产业。[4]

第六部分至第八部分讨论存在学习曲线时的（次优的）福利经济学。在第六部分，我们将会考察一种跨时合同。如果这些合同是可以执行的，那么它们就会使自由进入的产业维持在社会最优水平。我们将会确定这些跨时合同。同时，我们还会指出，如果有很少的进入成本，那么所有分析都不能成立。

于是，这就会促使我们去考察政府管制的影响（第七部分和第八部分）。干中学以极端的方式引发了一种标准的冲突：对于整个社会而言，不受约束的垄断是不合意的，但是寡头垄断也许更差。原因在于，寡头垄断的市场结构下，在分享市场的过程中，各企业学习较少，从而产品价格就较高。我们的分析表明，对学习效应很强的产业中应用自由竞争原则，本身就是不合适的。在第七部分中，我们将会看到，对自由竞争原则的几种应用都表明了这一点。[5] 在第八部分，我们将在几种不同的条件下研究，当存在外国的竞争时，那几种广为人知的政府政策会产生哪些影响。我们将会正式地阐述幼稚产业论的一种说法，并且确定在哪些条件下短期的进口禁令会导致国际的帕累托改善（第八部分第1点）。最后，在第八部分第2点，我们指出，如果外国的产业有很强的学习效应，并且比本国产业的学习潜力大，同时假如本国的生产成本较高，而且本国居民对外国产品的需求占外国产品的很大一部分，那么就应该采取短期的进口补贴。这些补贴会促使外国企业生产更多的产品，从而更好地发挥其学习效应。本国进而会从未来更低的进口价格中获得本国利益的改善。

尽管干中学非常普遍，但是研究其经济含义的理论文献仍然很少。[6] 阿罗（Arrow，1962）以及卡尔多和莫里斯（Kaldor and Mirrlees，1962）以干中学的不同形式构建出了封闭经济的总量增长模型。这些文献都假定经济是竞争性的。然而，下面我们将会注意到（第五部分），当且仅当所有企业都能在瞬间，以无成本的方式从其他企业的经验中学到所有知识时，该假设才能成立。也就是说，当且仅当学习的溢出是完全的，经济才会是竞

---

4 我们还要强调，当预期到进入之后的市场竞争很激烈时，即使很少的进入成本，也能使竞争实际上不会发生的事实，是一个与干中学无关的一般性的结论（Dasgupta and Stiglitz，1987）。

5 对美国和西欧反托拉斯政策的说明，尤其是对隐含在《美国谢尔曼法案》中的自由竞争原则的说明，参见 Scherer（1980），chapter 19，20。

6 对干中学所引起的问题的一般性讨论，参见 Scherer（1980），pp. 250－252。

争性的。[7]这一极端的假设无法经受严格的推敲，它并不符合现实。前面已经简略地说明了，我们的实证分析就是为了研究学习现象对市场结构的影响。为了达到这一目标，我们将会在第三部分和第四部分第1点依次研究，学习效应对社会统一管理的产业的影响，以及它对私人垄断产业的影响。在第九部分，我们会列举本文的结论。下面以术语和符号等来开始我们的分析（第二部分）。

二、预备知识

有 n 家企业（或生产单位）生产一种同质的商品。将企业（或生产单位）标示为 i，j（$=1, 2, \cdots, n$）。下面的分析主要关注两个生产期间，$t=0, 1$。不过，有效的约束并不是“两个生产期间”，而是“有限多个生产期间”。其原因在于，我们将会注意到，很容易就能将分析结果推广到有限多个生产期间。我们希望能将一个产业发展的早期（幼稚阶段），与它的后期（成熟阶段）区分开来。在产业发展的早期，学习现象是非常明显的；而在产业发展的后期，学习效应就不存在了。

为了集中考察我们所要研究的产业，假设收入效应可以忽略不计。在每一期，对该商品的市场需求都由向下倾斜的需求曲线 $p(Q)$ 给定，其中 Q 是总产出，p 是商品的销售价格（假设需求的变动不能带来任何收益）。进一步地，我们假设当 $Q\to\infty$ 时，就有 $p(Q)\to 0$。假设 $p(Q)Q$ 是严格凹的。也就是说，产业的边际收益是总产出的减函数。

令 c_0^i 为企业 i 在第一期的单位生产成本。为了简化分析，我们假设，对于所有的 i，都有 $p(0) > c_0^i > p(\infty) = 0$。学习是以生产经验为基础的。将企业 i 在第二期（$t=1$）的单位生产成本记为 c_1^i，我们就可以得到

$$c_1^i = c_1^i\left(Q_0^i + \alpha \sum_{j\neq i} Q_0^j\right)$$

其中，Q_0^i（相应地，Q_0^j）为企业 i（相应地，j）在第一期的产出。在此，α 是一个常数，它表示学习溢出的水平，而且有 $0\leqslant\alpha\leqslant 1$。由于存在学习效应，所以我们假设 $c_1^i(\cdot)$ 是一个单调递减的函数，而且 $c_0^i(0) = c_0^i$（原文如此，疑有误，应为 $c_1^i(0) = c_0^i$。——译者注）。若 $\alpha=1$，则学习

7 事实上，即使学习是针对具体企业而言的，一个有着大量进行数量竞争的相同企业的产业，在对称的古诺均衡状态下，它大致会是竞争性的。但是，在这一情形中，没有企业会学到很多东西。原因在于，在每一期各企业的生产量都可以忽略不计。

存在于整个产业，而且溢出是完全的。此时，与阿罗（Arrow, 1962）、卡尔多和莫里斯（Kaldor and Mirrlees, 1962）以及斯托基（Stokey, 1986）等人的假设是一致的。这是一种相当极端情形。另一极端情形是 $\alpha=0$。这就意味着，学习是视具体企业而言的。这种情形没有太大的问题。可以推知，只要 $\alpha<1$，我们的分析结论就不会随 α 的变化而变化。

三、国有企业与对边际成本定价的偏离

1. 完全最优的定价

假设企业是国有的。首先假定可以用一般税收来弥补损失。我们令 $u(Q)=\int_0^Q p(\bar{Q})\mathrm{d}\bar{Q}$，并且为了不失一般性，假设社会贴现率为零。为简化分析，假设所有潜在的生产单位都是相同的。如果溢出是不完全的（即 $\alpha<1$），显然就应该只有一个生产单位能从学习上获利。假如 $\alpha=1$，那么生产单位数目的多少就是无关紧要的。从而，我们就假设只有一个单位进行生产。

因此，问题就是选择 Q_0 和 Q_1（分别是 $t=0$ 和 $t=1$ 时的产出水平），以便实现下式的最大化

$$u(Q_0)-c_0Q_0+u(Q_1)-c_1(Q_0)Q_1 \tag{1}$$

下面我们将假定，这个最优化问题和第四部分中垄断企业的最优化问题，都有内点解。而要得到内点解，一个充分条件就是 $c_1(\infty)>0$。假定此充分条件得到满足，那么式（1）取得最大值的必要条件就是

$$p(Q_0)=c_0+c'_1(Q_0)Q_1 \tag{2a}$$

和

$$p(Q_1)=c_1(Q_0)^{8} \tag{2b}$$

式（2b）表示的是在成熟阶段企业的最优定价政策。此时，没有学习效应。因此，这就是边际成本定价法则。但是，在幼稚阶段，存在学习现象。而且，学习就蕴涵着一种跨期外部性。式（2a）就表示怎样才能将这种跨期外部性内部化。它表明只要在该企业中存在学习的空间，那么产品价格就会低于当时的边际成本。从而以当时的价值看，该企业就会遭受一定的损失。可以用一般税收来弥补这些最优的损失。

---

8 一撇表示导数。

2. 收入约束下的最优定价

假定该企业必须要在规划期限内自行弥补其损失。具体而言，就是要求它“平衡自身的预算”。因此问题就是在下式的约束下选择 Q_0^i 和 Q_0^j 去最大化式（1），

$$p(Q_0)Q_0 - c_0Q_0 + p(Q_1)Q_1 - c_1(Q_0)Q_1 \geqslant 0 \tag{3}$$

令 λ 为与式（3）有关的拉格朗日乘数。于是，设 $m(Q) \equiv p'(Q)Q + p(Q)$（$m(Q)$ 为边际收益），我们就可以将最优化的必要条件写成：

$$p(Q_0) - c_0 - c'_1(Q_0)Q_1 = -\lambda[m(Q_0) - c_0 - c'_1(Q_0)Q_1] \tag{4a}$$

和

$$p(Q_1) - c_1(Q_0) = -\lambda[m(Q_1) - c_1(Q_0)] \tag{4b}$$

或者等价地写成：

$$p(Q_0)[1 - \lambda\epsilon(Q_0)/(1+\lambda)] = c_0 + c'_1(Q_0)Q_1 \tag{5a}$$

和

$$p(Q_1)[1 - \lambda\epsilon(Q_1)/(1+\lambda)] = c_1(Q_0) \tag{5b}$$

其中，$\epsilon(Q) = -p'(Q)Q/p(Q)$，它是需求弹性的倒数。

如果最优化问题（1）是一个凹规划，那么约束条件（3）就是紧的。此时，式（4a）至式（4b）［和等价式（5a）至式（5b）］就是拉姆齐规则，表示对边际成本定价的最优偏离。假如 $\lambda > 0$，我们立刻就可以从式（4b）中推出 $p(Q_1) > c_1(Q_0)$；同时，可以从式（3）（它实际上是一个等式）中推出 $p(Q_0) < c_0$。换而言之，当存在学习效应时，就需在成熟阶段将商品价格定在边际成本之上，以弥补幼稚阶段的损失（式（4a））。

一个有意思的问题是，收入约束式（3）会对两个时期的最优产出水平产生怎么样的影响。假设 $c_1(Q_0)$ 充分地凸，以至于能使式（1）为凹。于是，就可以看到，如果在幼稚阶段次优的产出低于最优的产出，那么在成熟阶段也一样如此。实际上，根据式（2a）至式（2b）以及式（4a）至式（4b）很容易就可以发现，在 $t=0$ 时，次优的产出确实要低于最优的产出。因此，我们就有

命题1：如果对国有企业施加一个收入约束，那么就会使每期的产出水平都下降。

图1用图形的方式表达了命题1的内容。实线 $Q_0(Q_1)$ 是式（2a）的解，位于它下面的那条虚线是式（5a）的解。它们分别是不存在和存在收入约束时 Q_0 的最优值。而且，Q_0 是 Q_1 的函数。同样，可以认为式（2b）和式（5b）确定了对应于 Q_0 的每个 Q_1 的最优值。在图1中它们分别由实

线 $Q_1(Q_0)$ 以及它下面那条虚线来表示。Q_0 越高，则下一期的边际生产成本就越低，从而 Q_1 就越大。由于收入约束意味着，Q_0 和 Q_1 的每个值都会下降，所以对应的 Q_1 和 Q_0 的值就会下降，于是两条虚线就都位于相应的实线之下。

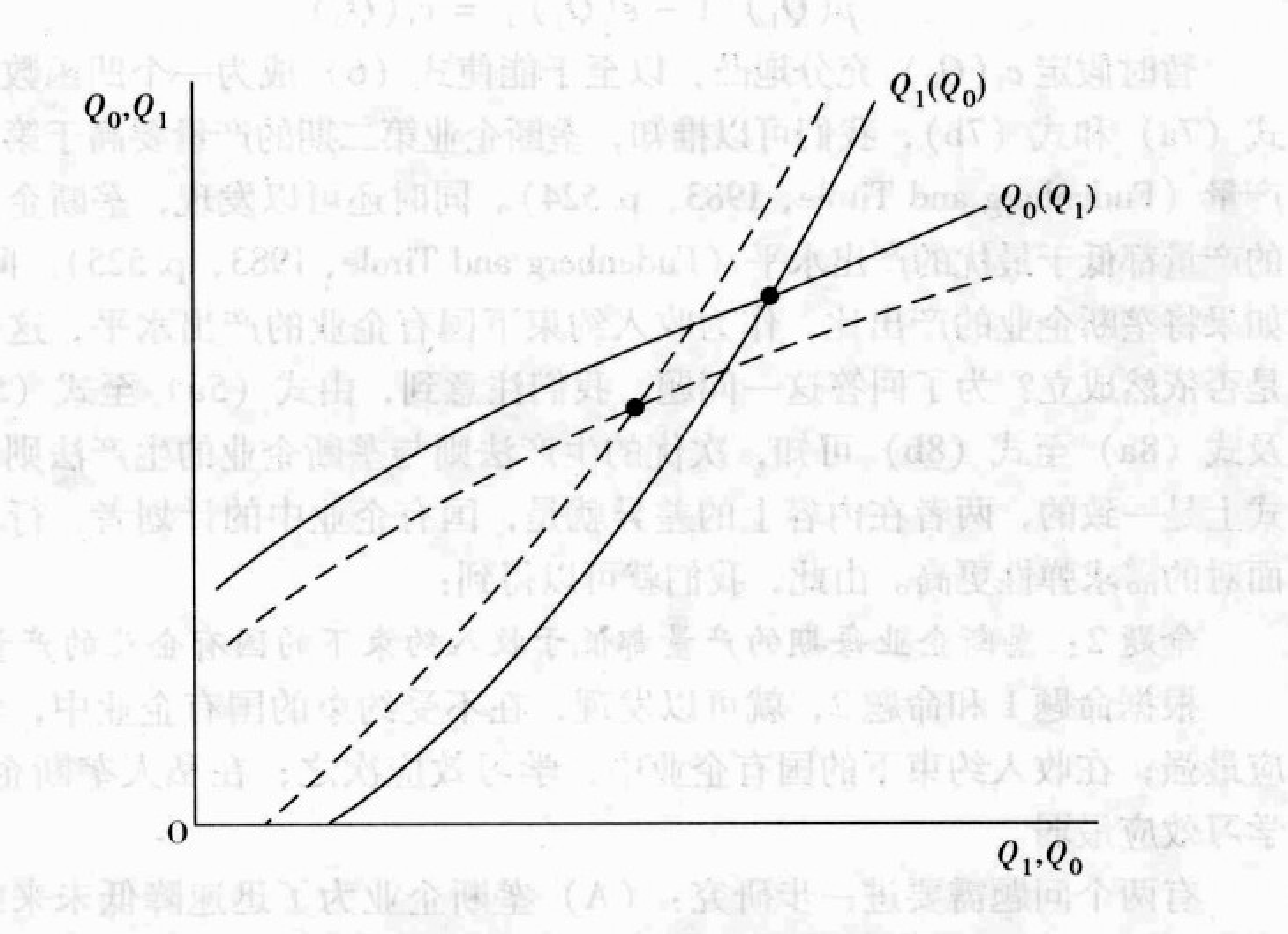


图 1　存在和不存在收入约束时国有企业的产出水平

四、寡头垄断企业之间的古诺竞争与垄断的形成

现在，我们转而考察一个私人企业。本部分第 1 点的主题是分析完全垄断这一最简单的情形。为了后面研究的方便，我们需要这一分析。在本部分第 2 点，我们将会分析展开古诺竞争的寡头垄断企业。

1. 垄断

我们考虑一个垄断企业，它不用面对竞争对手的进入威胁。假设资本市场是完善的，并且（不失一般性）假设利息率为零。与我们在分析国有企业时指出的原因一样，可以假设垄断企业只选择一个生产单位。因此，问题就是选择 Q_0 和 Q_1 以使下式最大化

$$p(Q_0)Q_0 - c_0Q_0 + p(Q_1)Q_1 - c_1(Q_0)Q_1 \tag{6}$$

这样，就可以得到必要条件

$$m(Q_0) = c_0 + c_1'(Q_0)Q_1 \tag{7a}$$

$$m(Q_1)=c_1(Q_0) \tag{7b}$$

可以将它们等价地写成

$$p(Q_0)[1-\epsilon(Q_0)]=c_0+c_1'(Q_0)Q_1 \tag{8a}$$

和

$$p(Q_1)[1-\epsilon(Q_1)]=c_1(Q_0) \tag{8b}$$

暂时假定 $c_1(Q_0)$ 充分地凸，以至于能使式（6）成为一个凹函数。根据式（7a）和式（7b），我们可以推知，垄断企业第二期的产量要高于第一期的产量（Fudenberg and Tirole，1983，p. 524）。同时还可以发现，垄断企业每期的产量都低于最优的产出水平（Fudenberg and Tirole，1983，p. 525）。问题是，如果将垄断企业的产出比，作为收入约束下国有企业的产出水平，这个结论是否依然成立？为了回答这一问题，我们注意到，由式（5a）至式（5b）以及式（8a）至式（8b）可知，次优的生产法则与垄断企业的生产法则，在形式上是一致的，两者在内容上的差异就是，国有企业中的计划者，行动时所面对的需求弹性更高。由此，我们就可以得到：

命题 2：垄断企业每期的产量都低于收入约束下的国有企业的产量。

根据命题 1 和命题 2，就可以发现，在不受约束的国有企业中，学习效应最强；在收入约束下的国有企业中，学习效应次之；在私人垄断企业中，学习效应最弱。

有两个问题需要进一步研究：（A）垄断企业为了迅速降低未来的生产成本，它会一直将价格定在低于单位生产成本的水平吗？（B）垄断企业会一直使第二期的价格高于第一期的价格吗？对这两个问题的回答都为："是的"。而且，为了找出能使"是的"这个答案成立的条件，我们只需要将这些问题正式地表达出来。这样也有助于展开后面的分析。

关于问题（A），我们定义

$$V(Q_0)\equiv\max_{Q_1}\{p(Q_1)Q_1-c_1(Q_0)Q_1\} \tag{9}$$

令 $\tilde{Q}_1(Q_0)$ 为式（9）的解。于是，就有

$$V'(Q_0)=-c_1'(Q_0)\tilde{Q}_1(Q_0)$$

现在令 $\hat{Q}_0$ 为 $p(Q_0)=c_0$ 的解。当产量为 $\hat{Q}_0$ 时，由于产量的边际增加，而带来的当期利润的增加量就是

$$m(\hat{Q}_0)-c_0=-c_0\epsilon(\hat{Q}_0)$$

而且第二期利润的增加量为 $V'(\hat{Q}_0)$。我们可以推知，当下式成立时，垄断

企业将会把第一期的价格，定在低于单位成本的水平上

$$V'(\hat{Q}_0) = -c_1'(\hat{Q}_0)\tilde{Q}_1(\hat{Q}_0) > c_0\epsilon(\hat{Q}_0) \tag{10}$$

而且，当下式成立时，垄断企业就会把第一期的价格，定在高于单位成本的水平上，

$$V'(\hat{Q}_0) = -c_1'(\hat{Q}_0)\tilde{Q}_1(\hat{Q}_0) < c_0\epsilon(\hat{Q}_0) \tag{11}$$

更容易将问题（B）形式化地表达出来。由式（7a），我们可以发现，

$$m(Q_0) = c_0 + c_1'(Q_0)\tilde{Q}_1(Q_0) \tag{12}$$

这样，我们就可以得到第一期的产出 Q_0^m。换而言之，第二期的价格高于或低于第一期的生产成本的条件是

$$p(Q_1(Q_0^m)) > c_0, \tag{13a}$$

或

$$p(\tilde{Q}_1(Q_0^m)) < c_0, \tag{13b}$$

2. *古诺双头垄断*

我们假设市场会持续 $T+1$ 期，其中 T 是一个有限的但足够大的数。由于我们要分析的是数量竞争，所以最好是假设这个产业是双头垄断产业。我们可以很方便地将这里的分析推广到寡头垄断的情形中去。为了获得显性解，我们假设每期的市场需求曲线都是线性的，即

$$p(Q) = A - BQ; \quad A, B > 0 \tag{14}$$

进一步地，我们假设，只要生产成本为正，那么学习曲线就是线性的，即

$$c_t^i = \max\{0, c_0^i - \beta\sum_{\tau=0}^{t-1} Q_\tau^i\}, \text{对于 } i = 1,2, T \geqslant t \geqslant 1, \text{其中 } \beta > 0 \tag{15}$$

式（15）表示的学习曲线有两个参数 c_0^i 和 β。前者度量的是学习所能发挥的作用的限度，而后者衡量的则是学习的速度。式（15）非常清楚地表明，经验积累到一定程度后，学习就会停止。一旦单位成本下降为零，则学习的收益就会迅速递减。

假设有一家企业（比如企业 1）与其竞争对手相比，开始时就拥有成本优势。我们想研究各企业不采取合作行动时的产业演化。首先假定 T 很大，而且各企业会以很低的贴现率去贴现未来的利润。显然就会有，在均衡状态，企业间的差距在长期将会消失。原因在于，即便是生产有限的产品，劣势企业（企业 2）也能将其生产成本降低至零。从这一时点开始，各企业就会获得双头垄断利润。假如时间很长，而且贴现率很低，那么双头垄断利润

的现值就会很大。很明显，此时所有企业都能从生产中获利。每家企业的市场份额在长期看来是相等的。

但是，如果贴现率很大，那又会怎样？上面的论述不再成立，而且劣势企业会在它与竞争对手之间的成本差异变大时，逐渐退出市场。下面我们分析这一选择是如何发生的。

假如贴现率很大，那么每个企业都会非常看重当期利润。假定所有企业展开数量竞争，优势企业（企业 1）第一期的产量就会高于其竞争对手的产量。由于学习曲线是线形的，所以企业 1 的成本减少量就大于企业 2 的成本减少量。因此，劣势企业仍然处于劣势地位，实际上，处境更差，并如此下去。于是，问题就是，确定那些能导致产业集中度提高的条件。我们通过研究一种特别的情形来达到这一目标。在此情形中，所有企业都非常短视。也就是说，所有企业都以一个无穷大的贴现率来贴现未来的利润。

令 Q_t^i 表示企业 i（$i=1$，2）在 t 期的（古诺）均衡产出，并假设它为正。于是，利用式（14），通过常规的计算就可以得到：

$$Q_t^1 = [A + c_t^2 - 2c_t^1]/3B, \text{和} \quad Q_t^2 = [A + c_t^1 - 2c_t^2]/3B \tag{16}$$

由式（15），我们得知

$$c_t^i = \max[c_{t-1}^i - \beta Q_{t-1}^i, 0] \tag{17}$$

假定 c_t^i 为正，并进一步假设 $t+1$ 期的均衡产出水平也为正。那么，很容易就可以得到

$$\begin{aligned} Q_{t+1}^1 &= [9B^2 Q_t^1 + \beta A - 5\beta c_t^1 + 4\beta c_t^2]/9B^2 \\ Q_{t+1}^2 &= [9B^2 Q_t^2 + \beta A - 5\beta c_t^2 + 4\beta c_t^1]/9B^2 \end{aligned} \tag{18}$$

由式（16）可知，如果假设 $c_t^1 < c_t^2$，则 $Q_t^1 > Q_t^2$，而且式（16）中的 Q_t^1 和 Q_t^2 都为正。于是，根据式（18）可知，若 $c_t^1 < c_t^2$，则

$$Q_{t+1}^1 / Q_{t+1}^2 > Q_t^1 / Q_t^2 \tag{19}$$

换而言之，只要企业间存在成本差异，那么产业集中度就会提高。

上述分析假定，全部企业在所有期间都进行生产。当劣势企业退出产业时，产业集中度能提高至无穷大的水平吗？下面将阐述这一问题。

令 $\Delta = c_0^2 - c_0^1 > 0$ 表示最初的成本差异。由于两家企业第一期的产量为式（16）中的产量，所以就有 $A + c_0^1 > 2c_0^2$。换句话说，就是

$$A - c_0^1 > 2\Delta \tag{20}$$

将式（16）代入式（18），我们就可以得到

$$9B^2Q_1^2 = 3B[A + c_0^1 - 2c_0^2] + \beta[A + 4c_0^1 - 5c_0^2] \tag{21}$$

现在我们假设 $A - c_0^1 < 5\Delta$，这与式（20）是完全一致的。于是，式（21）的右边第一个括号中的表达式的值为正，而第二个括号中的表达式的值为负。若 β "很大"，则式（21）中的 Q_1^2 就非正。不过，这就是劣势企业退出产业的另一种说法：当第二期的成本差异很大时，劣势企业就会停止生产。因此，我们就可以推出：

命题 3：当各企业都以很高的贴现率去贴现未来的利润，而且学习的空间又大致保持不变时，只要学习没有停止，产业集中度就会保持增长的态势。如果学习的速度足够快，那么在一定条件下，该产业在长期就会变为垄断产业。

图 2 和图 3 表明了这些可能的结果。为了方便起见，我们假设图 2 和图 3 中的时间是连续的。在每个图中，都假定企业 1 与其竞争对手相比拥有初始成本优势。在图 2 中，到企业 1 停止学习时为止，产业集中度一直都会提高。接着，企业 2 就能逐渐提高市场份额（即产业集中度会逐渐下降），直至它停止学习。从企业 2 停止学习这一时期开始，所有企业的产量都保持不变，并平分市场份额。

图 3 描绘了一种不同的情形。在此图中，产业集中度会持续提高，直至最终形成垄断。在有限的时间内，成本差异变得足够大，从而迫使企业 2 退出市场。

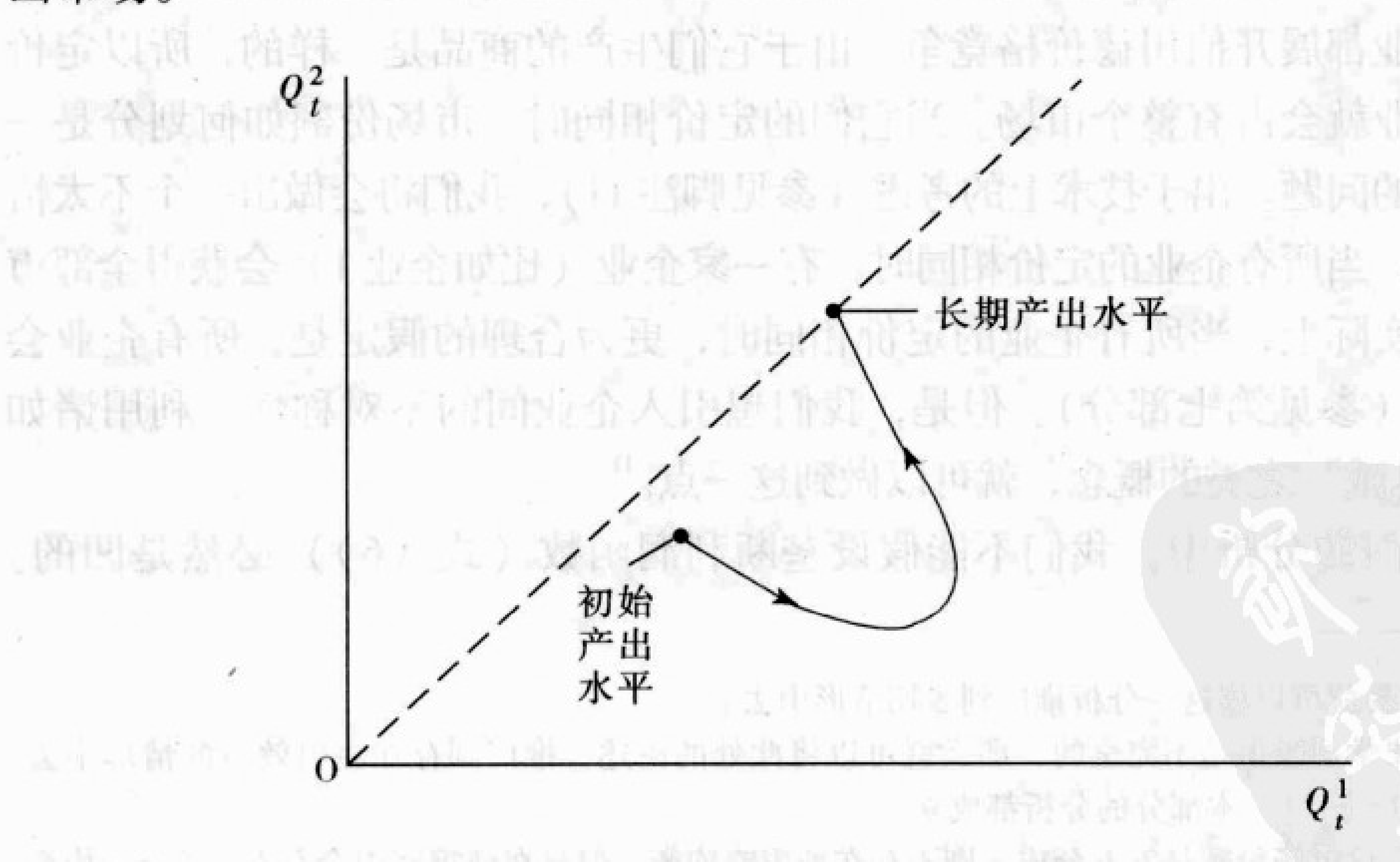


图 2　市场份额的演变 1：双头垄断的持续

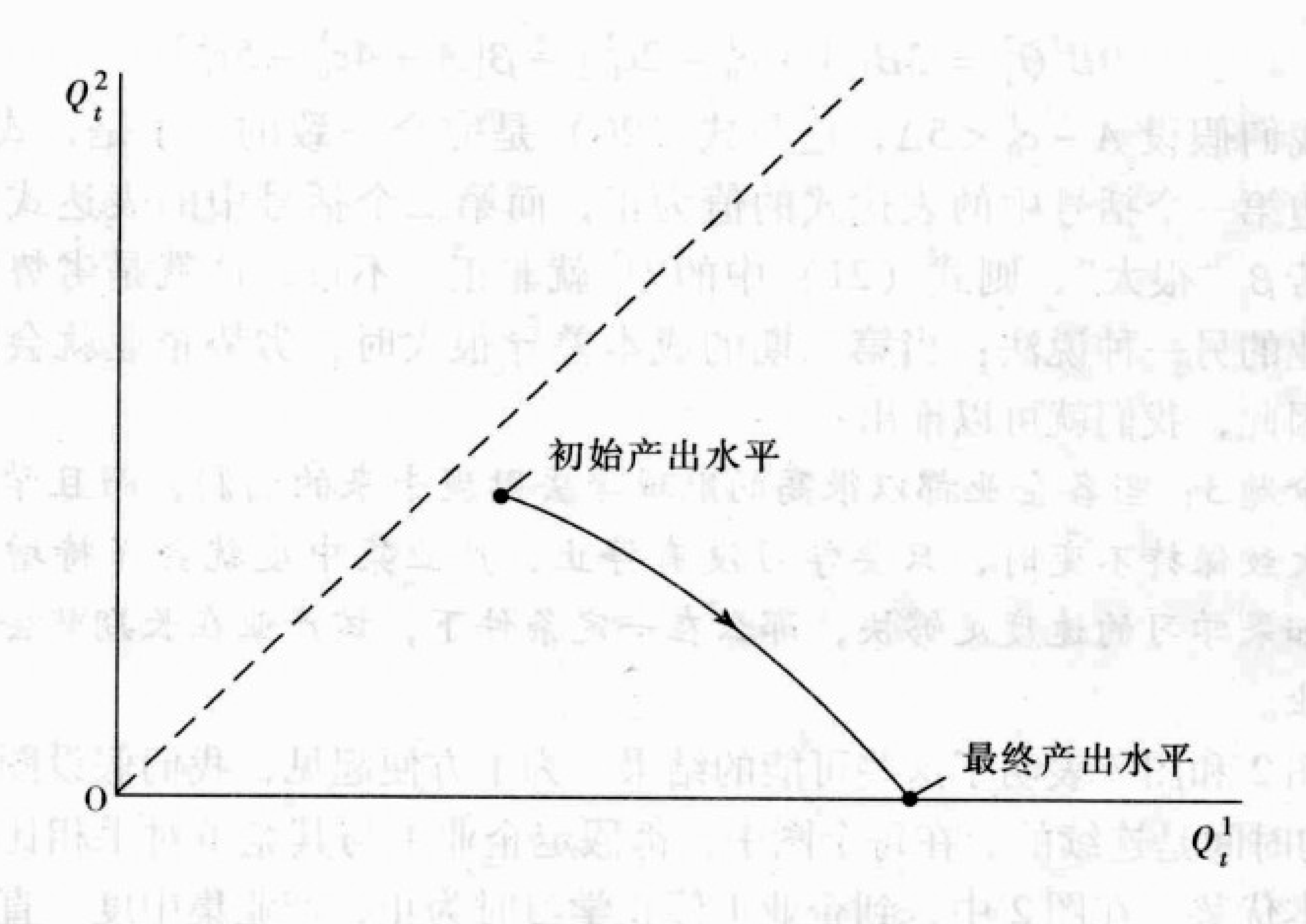


图3 市场份额的演变2：垄断的形成

五、寡头垄断企业之间的伯川德竞争

1. 外生的市场结构

产业中有两个企业，每家企业的单位生产成本最初都为 c_0。生产期间分为两期 $t=0$ 和 $t=1$。[9]为了简化分析，我们假设学习是视具体企业而言的。[10]在每一期所有企业都展开伯川德价格竞争。由于它们生产的商品是一样的，所以定价较低的企业就会占有整个市场。当它们的定价相同时，市场份额如何划分是一个很微妙的问题。出于技术上的考虑（参见脚注11），我们将会做出一个不太恰当的假设：当所有企业的定价相同时，有一家企业（比如企业1）会获得全部市场份额。实际上，当所有企业的定价相同时，更为合理的假定是，所有企业会平分市场（参见第七部分）。但是，我们想引入企业间的不对称性。利用诸如"消费者忠诚"之类的概念，就可以做到这一点。[11]

在下面的分析中，我们不能假设垄断利润函数（式（6））必然是凹的。

---

9 很容易就可以将这一分析推广到多期情形中去。

10 只要学习溢出是不完全的，那么就可以将此处的论述，推广到存在溢出效应的情形中去。因此，只要 $0 \leqslant \alpha < 1$，本部分的分析都成立。

11 若假设市场份额是五五分成，则不存在纯策略均衡。但是在纯策略中会存在一个 ϵ – 均衡。实际上，我们在本部分中算出的非对称分享规则下的均衡，大致等同于对称性分享规则下的 ϵ – 均衡（参见第六部分）。

于是，为了表征这一市场中的（子博弈精练，subgame perfect）均衡，我们定义

$$Q_1^*(Q_0) \equiv \arg\max_{Q_1} [p(Q_1) - c_1(Q_0) Q_1], \quad \text{其中 } p(Q_1) \leqslant c_0 {}^{12} \tag{22}$$

接下来，设 Q_0^* 为下式表示的零利润条件的最优解

$$p(Q_0)Q_0 - c_0Q_0 + p(Q_1^*(Q_0))Q_1^*(Q_0) - c_1(Q_0)Q_1^*(Q_0) = 0^{13} \tag{23}$$

现在，立即就可以得到：

命题4：双头垄断市场中只有一个子博弈精练均衡。在均衡状态下，优势企业（企业1）在 $t=0$ 和 $t=1$ 时，将其产品的价格分别定为 $P(Q_0^*)$ 和 $p(Q_1^*(Q_0^*))$，并占领整个市场。其竞争对手在两个时期的定价分别为 $p(Q_0^*)$ 和 c_0，并且不进行生产。

来自竞争对手的竞争使得优势企业的最大化利润变为零。但是，我们注意到式（22）和式（23）就隐含着 $p(Q_0^*) < c_0$。换而言之，在幼稚阶段，优势企业也不得不承担一定的损失。如果式（22）中的价格约束不是紧的，那么在成熟阶段，生产者就能获得垄断利润。

2. 潜在的竞争

潜在的竞争，会对现有企业的行为产生哪些影响？要分析这一问题，我们需要改变前面的模型结构，并假设在双头博弈展开之前，产业中并没有竞争对手。根据定义，此时产业中只有现有企业进行生产。竞争对手的进入是否重要是可以推断出来的。在第一期，现有企业的单位生产成本为 c_0。假如竞争对手在第一期就进入，那么它也能在单位成本 c_0 上进行生产。如果它不进入，根据定义，它就不能在市场上与现有企业展开竞争。与以前一样，学习是视具体企业而言的。

假设可以自由进入，但是禁止退出（读者很容易就能计算出可以无成本退出的结果）。此外，如果竞争对手在第一期没有进入，那么它就可以选择是否在第二期进入（当然，它也可以选择一直都不进入）。各方的行动如下。首先，竞争对手决定是否进入。如果决定进入，那么两个企业就展开价

12 因为我们假设当期的（总）收益对产出是严格凹的，所以 $Q^*(Q_0)$ 是一个函数，而不是一个对应关系。

13 由于式（6）并不一定就是 Q_0 和 Q_1 的凹函数，所以式（23）也许有多个解。

格竞争，并且在每一期都共享市场需求（我们假设在任何一期，当两个企业的定价相同时，现有企业都会占领整个市场。我们在此引入这一不对称性的动机，与本部分第1点是一样的）。另一方面，如果竞争对手不进入市场，现有企业就像一个垄断企业那样自由行事。然而，竞争者可以选择在第二期开始进入。如果选择进入，那么就展开价格竞争（但是，在这一情形中，竞争者要面对成本上的劣势。原因在于，它的单位生产成本仍然为c_0）。[14]假若竞争者在第二期仍然不进入，那么现有企业就再一次像垄断企业那样自由行事。

与命题4的证明一样简单，我们可以推知：

命题5：如果可以自由进入，那么市场中就存在三个子博弈精练均衡：

（ⅰ）立刻进入均衡，即竞争者立即进入市场；

（ⅱ）延迟进入均衡，即竞争者到第二期开始时，才进入市场；

（ⅲ）不进入均衡。

在任何一个均衡中，竞争者都不进行生产，也无法获得收益。现有企业会对三个均衡进行排序，它最偏爱的是（ⅲ），最不喜欢的是（ⅰ）。在均衡（ⅰ）中，它得不到任何收益。

现在，我们就开始描述均衡状态，以及各个企业在每一个均衡中的最终利润。为了达到这一目标，我们定义：

$$Q_0^{**} \equiv \max[\arg\max_{Q_0}[p(Q_0)Q_0 - c_0 Q_0 + p(Q_1^*(Q_0))Q_1^*(Q_0) - c_1(Q_0)Q_1^*(Q_0)]] \tag{24}$$

其中，$Q_1^*(Q_0)$由式（22）确定。接下来，用Q_0^m和Q_1^m表示式（8a）和式（8b）的解，即两个时期的垄断产出水平。很容易就可以确定，在不进入均衡状态下，现有企业的策略就是，在第一期将产品价格定为$p(Q_0^m)$，并宣布如果竞争者不进入市场，它就将第二期的价格定为$p(Q_1^m)$，如果竞争者选择进入市场，那它就将其第二期的价格定为$p(Q_1^*(Q_0^m))$。显然，在不进入均衡状态下，竞争者不会进入市场，所以现有企业能获得垄断利润。

类似地，很容易就可以断定，在延迟进入均衡状态下，现有企业会将第一期的价格定为$p(Q_0^{**})$，并宣布如果竞争者进入市场，它就将第二期的价

14 巴特查亚（Bhattacharya，1984）分析了另外一种情形。在他所分析的情形中，潜在的进入者可以在进入之前，就进行投资以降低成本。

格定为$p(Q_1^*(Q_0^{**}))$，如果竞争者不进入市场，那它就将进行垄断定价。竞争者的策略是，在第二期开始时进入市场，并将价格定为c_0。最后，立刻进入均衡状态下的策略，与命题4所描述的策略是一样的。所以，同样就可以很容易地断定命题5的剩余部分都是正确的。[15]

假如生产时期多于两个，上述分析会受到哪些影响？会有一些影响，但不会太多。如果有T个时期，而且学习一直都不会停止，那么在自由进入的情况下，就有$T+1$个子博弈精练均衡。就像命题5所表示的那样，在每一个均衡状态下，竞争者都不能获得任何收益。现有企业会对这些均衡进行排序，它最偏爱的仍然是不进入均衡。

因此，我们就可以推知，潜在的竞争并不能确保现有企业的利润下降为零。然而，我们怀疑是否只需关注不进入均衡就可以了。如果我们假设进入是无成本的，上述分析就值得怀疑了。所以，我们假设竞争者要进入市场，就必须要负担一个（沉没）成本k（>0）。现在，我们很容易就能断定：

命题6：如果竞争者要负担进入成本，那么市场中就只有一个均衡。在均衡状态下，竞争者不会进入市场，现有企业像不受约束的垄断企业那样行事。[16]

命题6明确地指出，初始的成本优势在长期看来都具有决定性的作用。由于假定生产同质商品的企业会展开价格竞争，所以结论就显得非常清晰。如果商品不是同质的，那么就需要用一个非负的进入成本，来剔除所有其他均衡，而只保留一个均衡。在这个均衡中，竞争者一直都不会进入市场。

六、长期合同

在前面一部分的讨论中，我们得到两个主要结论。第一，如果竞争者进入市场后，企业间的价格竞争很激烈，那么进入成本就会使潜在的竞争，无法对现有企业构成威胁（进入之后的竞争越激烈，足以成为进入壁垒的进入成本就越低）。第二，立刻进入均衡的零利润结果也无法实现，它并不能

15 如果式（13b）成立，那么就利润而言，延迟进入均衡和不进入均衡，对现有企业都是一样的。原因在于，在这一情形中，当$Q_0^m=Q_0$时，式（14）中的约束就不成立。另一方面，如果式（13a）成立，那么这一论断就不成立。此时，延迟进入均衡中，现有企业获得的利润更少。

16 更正式地说，利用一个小的进入成本，就可以剔除那些实际上不可行的均衡。在我们目前的模型中，考虑竞争者的均衡战略，它可以剔除立刻进入均衡和延迟进入均衡。它的作用与在位者的某种类型的“颤抖的手”（trembling－hand）所起的作用相同。

维持第三部分第2点中的约束条件下的最优状态。现在我们继续考虑第二个结论，并将会看到，使立刻进入均衡无法达到收入约束下的最优状态的因素，就是长期合同的缺位。在下一部分中，我们将会讨论第一个结论的政策含义。下面，我们就转而考察一个两期模型。

与本部分第1点一样，我们假设所有企业都已经在市场上。接着，假设所有消费者都可以签订长期合同。因此，企业可以向消费者提供各时期的报价单。消费者将同意（这是有约束力的）在第一期购买许多生产单位的产品，在第二期也购买许多生产单位的产品。从理论上讲，提供最佳合同的企业应该能吸引所有的消费者。现在，我们开始设定博弈的过程。与前面一部分一样，我们假设所有企业都同时行动。在这一情形中，我们需要假设，其中有一家企业具有特权，它能在各企业提供相同的合同时，占领整个市场。或者我们假设，其中一家企业并不是一个追求利润最大化的企业，只要处于最大利润的邻域之中，它就感到满足了（否则的话，在纯策略中就不存在均衡）。另一种思路就是，遵循可竞争性理论的做法，让其中一个企业最先行动，另一个企业作为追随者展开行动。在下面的论述中，我们继续假设两个企业同时行动，并且让其中一个企业在势均力敌时占领整个市场。

假设所有消费者都是相同的，从而他们都会同意最佳的合同。如果企业能进行非线性定价，那么在均衡状态，第三部分第1点中的全局最优状态将可以得到维持。每个企业都会以式（2a）和式（2b）所确定的价格销售商品，另外还会收取一笔足够大的固定费用，以确保其利润恰好为零。如果非线性定价不可行，那么市场均衡将会维持在第三部分第2点中的收入约束下的最优状态。此时，企业会按照式（5a）和式（5b）进行拉姆齐定价。

这类长期合同通常并不存在。但是，若存在这类长期合同，则它们本身就是一种进入壁垒，可以阻止那些偶尔会开发出新的更好的生产技术，或者生产出质量更好的商品的企业进入市场。此外，我们不能忽视签订合同需要承担一定的（沉没）成本这一事实。而且，正如我们在本部分第3点所注意到的那样，这是一种重要的进入壁垒。于是，我们就回到了命题6：不管多小的进入成本，都是实际竞争的一个障碍。

对于一个存在干中学现象的产业而言，这一结论的政策含义是什么？现在我们转而考虑这一问题。

七、自由竞争原则

反托拉斯政策有很多种形式，但是它们都遵循一种普通的观念，即如果一个产业中只有一个大企业，那么从本质上讲这一定不是什么好事。除非学习上的溢出是完全的，否则这种观念就是有问题的。原因在于，当溢出是不完全的时候，或者每个企业的收益并不是充分递减的时候，该产业就是自然垄断产业。假如学习效应很明显，那么垄断并不一定就是最差的市场结构。对整个社会而言，双头垄断（或寡头垄断）带来的结果也许更差：这可能会延长产业的幼稚阶段。于是，我们就可以推知，那种为一个企业设定最高市场份额的政策，也许会损害社会福利。

上述分析表明，拆分一个垄断产业可能会损害社会福利。我们转而讨论另一种可能会降低社会福利水平的反托拉斯政策。也就是说，当一家企业将价格定在低于当期边际成本水平时，政府就认定这是掠夺性行为的做法，将会降低社会福利水平。我们已经在第三部分中注意到，即使不考虑进入因素，垄断企业仍然可能会将价格定在低于当期边际成本的水平上（条件式(10)）。问题的关键是，监管机构是否能获得足够的信息，将掠夺性定价与存在很强的学习效应时的垄断定价区别开来。如果监管机构做不到这一点，并且迫使垄断企业在第一期将价格就定为 c_0，那么只要有很小的进入成本，就仍然会阻止竞争者进入市场。此时，该产业仍然是垄断产业，垄断企业各期的利润都降低了，并且会使各期的价格都上升（在第二期也是这样，其原因在于，学习效应变小了）。因此，这一政策损害了所有人的利益。

八、贸易政策

1. 幼稚产业论

幼稚产业论在极大程度上都会假设，外国的生产者是竞争性的，特别是，在外国生产者中，已经不存在学习效应。所以令 $\bar{c}$ 为外国的单位生产成本。假定运输成本为零，于是商品的世界价格就是 $\bar{c}$。与前面一样，令 c_0 为国内生产者第一期的单位生产成本，$c_1(Q_0)$ 为初始产出为 Q_0 时国内生产者第二期的单位生产成本。与此相关的一个假设是，我们假定 $c_0>\bar{c}$，并且存在 $\hat{Q}>0$ 使得 $c_1(Q)<c$ 对于所有 $Q\geqslant\hat{Q}$ 都成立。我们用 $p_d(Q)$ 来表示每一期的国内需求，$p_w(Q)$ 表示每一期的世界需求。

我们假设只有一个本国企业（命题 6 为这一假设提供了合理的根据）。如果没有政府干预，本国企业会进行生产吗？即使在有竞争性的外国企业的

情况下，只要利润的现值为正，它就会进行生产。为了形式化地阐述这一观点，我们将最优化问题写成

$$\pi \equiv \max_{Q_0, Q_1}[p_w(Q_0)Q_0 - c_0Q_0 + p_w(Q_1)Q_1 - c_1(Q_0)Q_1],$$
$$\text{其中} \quad p_w(Q_0), p_w(Q_1) \leqslant \bar{c} \tag{25}$$

如果 $\pi > 0$，那么本国企业将肯定会进行生产，并可能会通过限制性定价而垄断整个世界市场。若 $\pi < 0$，它根本就不会进行生产。我们假设是后一种情形，那么此时需要政府干预吗？

下面，我们并不探讨最优的政府政策，而是讨论在哪些条件下，诸如进口禁令之类的暂时性保护措施是合意的。为了简化分析，我们只考虑用进口禁令来保护幼稚产业的情形。[17]

令 $p_d(Q) = \bar{c}$ 的解为 $\overline{Q}_d$。由于 $\pi < 0$，所以我们注意到，在自由放任的政策下，本国的福利 S 为

$$S \equiv 2[u_d(\overline{Q}_d) - \bar{c}\overline{Q}_d], \text{其中} \ u_d(Q) = \int_0^Q p_d(\tilde{Q})\,\mathrm{d}\tilde{Q} \tag{26}$$

如果对本国企业实施幼稚产业保护，那么当它为国内市场生产时，它就不受国际价格约束。不失一般性地，我们可以假定，在幼稚阶段参与国际市场竞争并不符合本国企业的利益。当然，在成熟阶段，它也许会参与国际竞争。那时，如果它参与国际竞争，它将会垄断全球市场。我们自然会考虑，它想在第二期垄断全球市场这种情形。于是，我们将最优化问题写成

$$\bar{\pi} \equiv \max_{Q_0, Q_1}[p_d(Q_0)Q_0 - c_0Q_0 + p_w(Q_1)Q_1 - c_1(Q_0)Q_1],$$
$$\text{其中} \ p_w(Q_1) \leqslant \bar{c} \tag{27}$$

应该要注意到即使 π 为负，$\bar{\pi}$ 仍然可能为正。由式（27）可知，如果 $\bar{\pi} > 0$，那么就可以通过暂时的进口禁令，来诱使本国企业进行生产。在这一情形中，本国企业在幼稚阶段只为国内市场生产，而到了成熟期就能垄断全球市场。

但是，与自由贸易相比，进口禁令能提高福利水平吗？它能提高福利水平。为了看清这一点，令 $\overline{Q}_0$ 和 $\overline{Q}_1$ 为式（27）表示的最优化问题的解，并考虑式（27）中的约束条件非紧（即 $p_w(\overline{Q}_1) < \bar{c}$）的情形。于是，这一保护的结果就是，在第二期国内的消费者也能获得收益，而且这也许能为贸易保护提供足够的理由。正式地说，在幼稚产业的保护期结束之后，本国企业

17　同时，我们还会研究在幼稚阶段，对本国的企业进行生产补贴的合意性。

在成熟期的销售量设为 $\overline{Q}_1^d \equiv p_d^{-1}(p_w(\overline{Q}_1))$。于是，在贸易保护下本国的福利 $\overline{S}$ 为

$$\overline{S} = u_d(\overline{Q}_0) - c_0\overline{Q}_0 + u_d(\overline{Q}_1^d) - c_1(\overline{Q}_0)\overline{Q}^d + [p_w(Q_1) - c_1(\overline{Q}_0)](\overline{Q} - \overline{Q}_1^d) \tag{28}$$

由于 $\overline{Q}_1^d > \overline{Q}_d$，所以在一定条件下就有 $\overline{S} > S$。也就是说，在一定条件下，暂时的进口禁令比自由贸易更有利于本国经济。同时，由于 $p_w(\overline{Q}_1) < \overline{c}$，所以在第二期国外的消费者也能从贸易保护政策中获利。因此，我们说，进口禁令可以带来帕累托改善。

所有这些分析都假定，本国的垄断企业不能与本国的消费者签订长期合同。如果能签订长期合同，那么就不需要对幼稚产业进行保护。原因就在于，此时幼稚产业可以自行发展。当某些原因使得这些长期合同无法执行时，就需要政府进行干预。在我们所举的例子中，消费者在第二期就能获得由于幼稚企业进入而带来的收益，但是幼稚企业却不能占有这些收益。因此，在自由贸易的条件下，幼稚企业的利润为负。而进口禁令会使消费者在第一期就为其第二期的收益做了支付。我们并没有说，进口禁令是次优的政策。我们只是找到了一些会使进口禁令优于自由贸易的条件。

2. 外国的学习与进口补贴

下面我们来考虑一种与标准的幼稚产业论很不一致的情形。有时候本国会面对一个外国的垄断企业，它仍然还有学习效应。此时，本国政府还应该鼓励本国企业进行生产吗？答案取决于许多因素。如果本国的产业同样也面临一条学习曲线，特别地，假如本国产业的学习效应更强，那么此处的结论与上一点的结论就是一样的。此外，对外国垄断企业的依赖也就意味着，本国市场上的利润被外国企业赚走了。与之相比较，我们知道，在国际双头垄断的市场结构中，所有企业的学习就变少了。这会对本国的福利产生影响。

考虑本国的生产成本很高这种情形。此时，就需要从外国的垄断企业那里进口商品。为了强调我们所要指出的结论，假设外国的垄断企业是国有企业。它的生产计划是以国家的福利为基础的，而不会考虑进口国的福利。从进口国的角度看，就会存在学习不足的问题。这就意味着需要采取进口补贴。如果学习效应足够强，并且进口国的市场充分大，那么对进口实行合理的补贴，就能诱使外国的垄断企业生产更多的产品，因此就会使未来的价格降低至合理的水平。在此水平上，进口国获得的收益恰好等于它以前所支付

的进口补贴。[18]

我们通常看到的是有差别的产品间贸易。正如我们已经看到的那样，学习效应可以用来说明，为什么一家从事生产的企业会是自然垄断企业。假如各国根据它在各种产品生产上的学习能力来确定其应该生产哪些商品，那么每个国家都会专门生产某些商品（很可能会超过一个商品）。上一段所概括的思想表明，各国都应该对其进口的产品进行额外补贴，这就像关税同盟中的互惠协定。

九、结论

干中学意味着，在生产中存在动态的规模经济。它与一种特定形式的沉没成本有关。幼稚阶段的生产也承担着部分投资功能。这种沉没成本是可以根据需要进行调整的。在本文中，我们分析了学习对一个产业的结构、行为和绩效的影响。除了一些具体的论证细节，我们可以将文章的主要结论总结如下：

（1）当学习效应很强时，就会出现主导企业，从而产业集中度就会提高。即便在竞争对手间的竞争近似于势均力敌的情况下，也会出现这种情况。

（2）如果市场是“可竞争的”，那么主导企业获得的利润现值就可以忽略不计。但是由于存在学习效应，时间就显得非常重要。主导企业早期所遭受的损失可以用后来的利润去弥补。因此，并不能用主导企业在成熟期获得的利润数据，去分析该产业的基本特征。

（3）只要有很小的进入成本，就可以保证产业是完全垄断。换而言之，这些进入成本可以使潜在的竞争完全失效。

（4）当学习效应很强时，自由竞争的原则是有问题的。将一个固定的市场份额确定为现有企业的市场份额的上限，这很可能是一种次优的选择。

（5）在学习阶段，即便是一个受保护的私人垄断企业，也希望将产品价格定在低于当期单位生产成本的水平。因此，如果现有企业的产品价格低于生产成本，这并不一定就意味着它在实施掠夺性定价。

（6）如果外国的学习曲线非常陡峭，并且本国的生产成本又很高，那么进口大国就会发现，实行进口补贴也许是符合自己利益的。

（7）确实有一些条件能使传统的幼稚产业保护论成立。换句话说，在这些条件下，进口禁令能比自由贸易带来更高的福利水平。

18　加特西斯（Gatsios，1987）详细地分析了，当存在生产的规模经济时，组建关税同盟的可能性和合意性。

参考文献

ALCHIAN, A. (1963), "Reliability of Progress Curves in Airframe Production", *Econometrica,* 31, pp. 679–693.

ARROW, K. J. (1962), "The Economic Implications of learning by Doing", *Review of Economic Studies,* 29, 155–173.

ARTHUR, W. B. (1983), "On Competing Technologies and Historical Small Events: The Dynamics of Choice under Increasing Returns", mimeo. Food Research Institute, Stanford University.

ARTHUR, W. B., ERMOLIEV, M. and KANIOVSKI, Yu M. (1986). "Path-dependent processes and the emergence of macro-structure", *European Journal of Operational Research.*

ASHER, H. (1956), *Cost-Quantity Relationships in the Airframe Industry,* The Rand Corporation (Santa Monica).

ATKINSON, A. B. and STIGLITZ, J. E. (1969), "A New View of Technological Change", *Economic Journal,* 59.

BALDWIN, R. (1969), "The Case Against Infant-Industry Tariff Protection", *Journal of Political Economy* 77, pp. 295–305.

BAUMOL, W. J., PANZER, J. C. and WILLIG, R. D. (1982), *Contestable Markets and the Theory of Industry Structure,* Harcourt Brace Jovanovich (New York).

BHATTACHARYA, G. (1984), "Learning and the Behaviour of Potential Entrants", *Rand Journal of Economics,* 15, pp. 281–289.

DASGUPTA, P. and STIGLITZ, J. (1987), "Potential Competition, Actual Competition and Economic Welfare", *European Economic Review,* forthcoming.

DAVID, P. A. (1985), "Clio and the Economics of QWERTY", *American Economic Review,* Papers and Proceedings, 75, pp. 332–337.

DAVID, P. A. (1987), "Some New Standards for the Economics of Standardization in the Information Age", in P. Dasgupta and P. Stoneman, Eds, *Economic Policy and Technological Performance,* Cambridge University Press.

DAWKINS, R. (1986) *The Blind Watchmaker,* Longman Scientific and Technical (Harlow).

FUDENBERG, D. and TIROLE, J. (1983), "Learning-by-Doing and Market Performance", *Bell Journal of Economics,* 14, pp. 522–530.

GATSIOS, K. (1987), "Learning-by-Doing and International Trade Policy", mimeo., University of Cambridge.

HABERLER, G. von (1987), "Liberal and Illiberal Development Policy", in G. M. Meier, ed., *Pioneers in Development,* Second Series, Oxford University Press.

HOLLANDER, S. (1965), *The Sources of Increased Efficiency: A Study of Dupont Rayon Plants,* MIT Press (Cambridge).

KALDOR, N. (1985), *Economics without Equilibrium,* University College, Cardiff Press.

KALDOR, N. and MIRRLEES, J. (1962), "A New Model of Economic Growth", *Review of Economic Studies,* 29, pp. 174–192.

LIEBERMAN, M. B. (1984), "The Learning Curve and Pricing in the Chemical Processing Industries", *Rand Journal of Economics,* 15(2), pp. 213–228.

MYRDAL, G. (1957), *Economic Theory and Underdeveloped Regimes,* Duckworth (London).

NEGISHI, T. (1972), *General Equilibrium Theory and International Trade,* North Holland (Amsterdam).

SCHERER, F. H. (1980), *Industrial Market Structure and Economic Performance,* Second Edition, Rand McNally (Chicago).

SPENCE, A. M. (1981), "The Learning Curve and Competition", *Bell Journal of Economics,* 12, pp. 49–70.

STOKEY, N. (1986), "The Dynamics of Industry-wide Learning", in W. P. Heller, R. M. Starr and D. A. Starrett, Eds, *Equilibrium Analysis: Essays in Honour of Kenneth J. Arrow, Vol II,* Cambridge University Press.

WAN, H. Y. and CLEMHOUT, S. (1970),"Learning-by-Doing and Infant Industry Protection, *Review of Economic Studies,* 37, pp. 33–56.

ZIMMERMAN, M. B. (1982), "Learning Effects and the Commercialization of New Energy Technologies: the Case of Nuclear Power", *Bell Journal of Economics,* 13, pp. 297–310.

福利经济学

福利经济学

经济学中一个最重要的问题是关于市场经济的效率：市场配置资源的效率如何，政府应该起什么作用。对于正在向市场经济转型的中国，这些问题极为重要。并且正如本文最后所列文献表明，这些问题在我的研究中处于重要地位。我特别为《斯蒂格利茨经济学文集》（以下简称《文集》）第三卷写作本文，以总结关键性的发现。

经济学中最重要并且最广为引用的也许是福利经济学基本定理——亚当·斯密的"看不见的手"理论——它断言竞争市场是有效率的。阿罗和德布鲁之所以获得诺贝尔奖，部分原因是他们严格证明了这个结果[1]，并说明了在何种意义上以及什么情形下，斯密是正确的。他们认为，竞争市场导致帕累托效率，就是说，在不使某些人处境变坏的条件下，没有人的处境可以变得更好。

这个结果形成了对市场经济信心的基础。如果这是普遍正确的，那么政府的作用就是有限的并且是明确的。当然，没有人期望非竞争性的市场是有效率的，也没有人期望存在公共物品和外部性（比如存在污染的情形）时经济是有效率的。但是这些"市场失灵"范围有限，而且人们很容易设计政府干预政策。政府可以用反托拉斯法保证市场是竞争性的；政府提供诸如国防的公共物品；政府也能借助市场机制（污染税或者许可权交易）处理像污染这样的外部性。一旦政府处理了这些市场失灵，我们就可以依靠市场力量来保证经济的（帕累托）效率。

一、阿罗—德布鲁与市场失灵理论

然而上述结论的成立需要一些高度限制性的条件——比如要求存在完备

1 Kenneth and Gerard Debreu（1954）；Gerard Debreu（1959），Arrew and Hahn（1971）.

的保险市场，并且，如果信息不是完美的，至少要求个人信念不受经济中发生的任何事情的影响。标准理论还假设技术是固定的，然而技术变化却是增长和发展的核心。

对于中国来说，当其经济向有特色的市场经济转型过程中，重要的是理解市场的局限性，从而理解什么是政府的合适角色。目前的许多争论——关于更自由的市场和政府起更大作用之间的争论——是源自自由市场的支持者没有正确地理解市场的局限性。本文的目的是回顾这场争论，以便人们更好地理解公共部门的适当角色。

在阿罗和德布鲁成功证明亚当·斯密"看不见的手"定理以后，很多人力图将他们的结论一般化。特别是，如果某些高度限制性——并且不现实——的假设被弱化，市场有效率的结论仍然成立吗？对这个问题的回答使我们更好地理解了市场的局限性：过去半个世纪的研究已经表明，阿罗和德布鲁发现市场有效率的假设条件是一个奇异集（singular set）。我们还知道一些简单的政府干预可以减轻市场失灵。这篇文章的很大部分讨论了为什么市场经常会失灵，以及政府可以做些什么。

凯恩斯和新古典综合理论

几十年以前，面对批评者的攻击，凯恩斯"拯救"了人们对市场经济的信念；在这以前，大萧条中的大范围失业破坏了这种信念，表明市场的运行不是有效率的。自由市场的支持者认为失业不会持续很长时间，但是随着大萧条的持续，市场的恢复力量（最多）很明显是微弱的。凯恩斯阐明了直接的政府行为——刺激经济的财政和货币政策——如何使经济恢复充分就业。萨缪尔森的新古典综合理论（1948）认为，一旦恢复充分就业，我们就可以重新依靠市场来保证效率。凯恩斯关于市场力量运行缓慢的断言无疑是正确的，正如《文集》第四卷中的论文所表明的，政府能够并且应该实行干预，保证充分就业和宏观经济稳定。对这一点，目前为止人们存在广泛共识。[2]然而，即便是萨缪尔森本

2 尽管这种共识是广泛的，但还是有异议。某些保守主义经济学家（比如已故的米尔顿·弗里德曼）认为，政府干预不仅一般来说没有效果，即便有效果也是反生产性的。例如，对货币政策的管理不善促成了大萧条。政府干预有时候不起作用，或者在错误的方向起作用——尽管在这一点弗里德曼是正确的，但是得出"一般都是如此"的结论则是错误的。自从政府开始遵循凯恩斯经济学的教义以后，西方经济表现出更长的扩张期和更短的衰退期。并且，随后的研究表明弗里德曼的理论和经验分析中存在漏洞。例如，当个人受信贷约束时，其消费不再仅仅取决于永久收入，有充分证据表明暂时性减税或失业救济有时候的确会显著地影响消费。

人出面阐述，[3] 我也没有被这些观点说服。为什么市场失灵只会以“大规模”的形式存在——失业的规模如此之大，以至于人们无法忽视它们？这更可能只是冰山的一角——它表明可能存在更广泛的市场失灵，即使每一种失灵是微小的，但是加总在一起可能会占 GDP 的很大一部分。如果市场有时候无法做出重大调整，造成四分之一的劳动力没有工作，我们可以想像市场会经常无法做出微小调整，无法有效配置资源。格林瓦尔德（Greenwald）和我提出了这种观点，并且描绘出它对于宏观经济分析的某些含义（Greenwald and Stiglitz，1986），中文译文见《文集》第四卷。[4]

批评竞争市场的基础假设

市场有效率的结果基于几个假设，我在一系列文章中对它们进行了质疑。第一个假设是企业最大化利润或股市价值。[5] 在我早期的工作中，我识别了许多企业行为看起来与价值最大化不一致的情形，比如，企业没有最小化税收负担（税后所得的损失数以十亿计[6]）[7]。我对公司治理的研究（在下文和《文集》第二卷有简要讨论）认为这些情形并非偶然。相反，由于信息不完美，经理有激励——和手段——以股东为代价中饱私囊。[8]

3　1963～1964 年我在 MIT 修了他讲授的研究生基本理论课程。后来，我给他做研究助理，编辑了他的 Collected Scientific Papers 的头两卷。他还是我的博士论文指导者之一。

4　我们也试图解释企业面对冲击时，为什么不会完全调整价格（或工资）。参见 Greenwald and Stiglitz（1988）。

5　正如我在下文指出的，当不存在每一日期和每一种可能性下证券的完全集合时，甚至价值最大化的真实含义都是不清楚的。参见我和山迪·格罗斯曼（Sandy Grossman）的论文“On Value Maximization and Alternative Objectives of the Firm，”（中文译文见本卷）。事实上，没有证券的完全集合，不同的持股人想要企业追求不同目标。特别请参见“Some Aspects of the Pure Theory of Corporate Finance：Bankruptcies and Take－Overs，”（1972c，中文译文见《文集》第二卷）；Grossman and Stighitz“Stockholder Unanimity in the Making of Production and Financial Decisions”。

6　这被称为“税收悖论”。最明显的是股息税悖论：某些方式可以将货币从公司转移到家庭部门，从而降低税收。这在 Stiglitz（1973），“Taxation，Corporate Financial Policy and the Cost of Capital”一文中第一次被讨论。在 Stiglitz（1982b）中讨论了其他悖论。在 Stiglitz（1987b）中，我对许多支付报酬的标准方法是否符合“理性”表示怀疑——可以设计一些支付方案，提供同样强的激励，而风险更小并且/或者税收更低。对于每一个问题，都有一大批的参考文献。

7　这种差别的原因是复杂的，并且与公司治理、经理面对的激励、经理和持股人之间的利益分歧有很大关系。参见 Stiglitz（1985）；Edlin and Stiglitz（1995）。

8　在 Edlin and Stiglitz（1995）中，我们证明经理如何制造信息不对称，阻碍竞争者收购企业的努力，并使他们自己有更多自由从企业抽取租金。更一般的观点是“代理”问题的存在，所有者不得不将决策权委托给经理，而经理通常不代表他们的利益。这最初在 Stiglitz（1974）（中文译文见《文集》第一卷）和 Ross（1973）中被提出。我还证明，那些机制（比如收购）——据说是用来“约束”没有最大化持股人利益的经理——常常是无效的。参见 Stiglitz（1972c）和 Stiglitz（1982b）。

另外一个假设是企业具有理性预期。尽管我对此怀有疑问——我找不到股市价值在大约一个季度以内（比如在 1987 年 10 月）急剧下降的理性预期解释——也只能让其他人从计量经济学角度系统地批评这一假设[9,10]。

我还发现竞争模型中的价格接受假设没有说服力。如果企业面临向下倾斜的需求曲线，它们就不会使价格等于边际成本；具备一定垄断力的企业会提高价格并将产出降到有效水平以下，行使这种市场权力。

当然，几乎没有一个部门是由一个企业控制的。但是在很多（如果不是大多数）部门，产品之间差别很大，企业面对向下倾斜的需求曲线。每一种新产品的生产都会有固定成本，问题是垄断竞争市场——企业面对向下倾斜的需求曲线——是否导致经济效率？迪克西特（Dixit）和我（1977）证明了在一种临界情形下，垄断竞争会带来经济效率，但一般来说不是这样，[11]（在某种意义上，这里的关键问题是市场是否在规模和多样性之间作了最佳权衡。由于建立新企业（生产更多种类产品）存在固定成本，生产更多种类产品只能以每个企业生产规模变小为代价；由于平均成本是递减的，这意味着生产每种商品的平均成本更高了。我们证明存在一种重要情形——参数建模中使用广泛的单位弹性需求曲线——市场进行了最优的权衡。但是一般来说，市场没有做出最优的权衡。[12]

我对于不完美信息的研究使我更加确信：对证明市场效率至为关键的价

9　Shiller 在《非理性的繁荣》（*Irrational Exuberance*），（2000）中提出，市场经常会出现过度波动。到目前为止，行为金融领域有一大批文献记录了理性和观察到的行为之间存在的差异。例如，参见 Tversky and Kahneman（1974）；Thaler（1990）；Kahneman（2002）。

10　然而，我证明了理性预期——即便是在竞争市场中——一般不能保证经济效率（参见下文）。在 Neary and Stiglitz（1983）中，我们还证明了理性预期意味着政府干预无效的观点（因为个人或者企业可能会采取抵消性行为）是根本错误。事实上，我们证明理性预期下，政府干预在刺激衰退经济方面比没有理性预期时更加有效。

11　Dixit and Stiglitz（1977）.

在 Stiglitz（2004）中我更详细地解释了不完美信息怎样产生不完全竞争。也可参见 Salop and Stiglitz（1987）以及 Stiglitz（1989a）中引用的其他论文。

12　即便那时，市场也只是在某种“受约束”的意义上最优，即政府不能一次性支付固定成本。我们指出某种基于信息的理由可以排除这种一次性支付。在 Stiglitz（1986）中，使用商品空间的很不相同的参数化，我证明了在更弱的假设下，市场是有效率的，但是要回答市场是否导致过多或者过少多样性的问题并不容易。在我探讨的特殊的参数化情形下，这取决于定义商品的特征空间维数大于还是小于 5。在另外一篇文章中，萨勒普（Salop）和我问了相似的问题，企业面临向下倾斜的需求曲线的原因是否与信息成本有关，S. Salop and Stiglitz，“In formation，Welfare and Product Diversity”，In *Arrow and the Foundations of the Theory of Economic Policy*，G. Feiwel（ed.），London：MacMillan，1987，pp. 328 – 340。

格接受的行为假设是不合适的。在完美信息下，企业相对于竞争对手轻微地降低价格，就会获取整个市场份额；在不完美信息下，它能获取的市场份额小得多，因为其他商店的消费者不一定知道企业降低了价格——企业再一次面临向下倾斜的需求曲线。尽管戴梦德（Diamond，1971）认为即便存在很小的搜寻成本，均衡价格也会是垄断价格；我已经证明，一般来说，均衡可以更好地用垄断竞争模型来描述也就是存在价格和工资稳定分布的模型。[13]但是关键问题还是一样的：无法保证市场实现效率的资源配置。

企业不仅可以利用信息不完美以及信息不完美带来的市场权力，企业还必须受到生产高质量产品的激励。在不完美信息下，企业可能生产低劣的产品，比如，这些商品不像初看上去那样经久耐用，它们的缺陷只有以后才能发现。企业会赢得生产优质商品的声誉，但是如果要它们维持这种声誉，就必须存在失去这种声誉的成本。这意味着在边际上，企业必须在乎销售下降带来的损失——价格要超过边际成本。然而标准竞争模型，是通过价格和边际成本相等来保证生产效率的，即生产一单位商品的边际社会收益（以价格度量）等于边际社会成本（生产的成本）。[14]

我对亚当·斯密"看不见的手"定理的最根本批评是这种观点背后的推理过程。它隐含地依赖于一个推断，如果经济中只存在一种扭曲，消除这种扭曲——使该部门中的价格等于边际成本——会是一种帕累托改善。如果经济中不止一种扭曲——当市场不完全或信息不完美时通常如此——则一般的次优定理[15]认为，消除一种扭曲不一定是帕累托改善，理由很简单：它可能使其余扭曲带来的资源错配情况更加恶化。

二、存在风险市场的最优性

在阿罗和德布鲁证明他们的经典结论以后，许多经济学家力图证明在更弱的条件下，市场经济也是有效率的。然而，大多数人的早期工作并没有研究我们刚才讨论过的根本问题——所有人都认为，如果存在不完美信息，市

13 比如，参见 Stiglitz（1985b）；Stiglitz（1987a）。

14 参见夏皮罗（Shapiro，1983）。夏皮罗和我将这个基本概念应用于劳动市场。在 Shapiro and Stiglitz（1984）中，证明了在均衡时，应该有持续的失业。我后来和雷（Patrick Rey）的工作（1996）证明了存在"代理"问题的市场均衡一般不是（受约束的）帕累托最优的，因为通过改变偷懒（shirking）的结果或者降低所生产产品的质量，企业行为会影响其他企业的激励。也请参见 Arnott and Stiglitz（1995）。

15 Lipsey and Lancaster（1956）.

场就不可能是有效率的。相反，大多数研究者致力于探讨阿罗和德布鲁提请人们注意的一组新的约束条件。阿罗和德布鲁假设一个完全的市场集合涵盖所有风险和所有未来日期的市场存在。比如，人们可以买卖 2008 年 3 月 15 日交货的苹果。显然，真实经济并不满足这个条件。但是市场经济的赞成者相信，必定存在某种情形使经济仍然有效率。

当然，缺乏某些风险市场的经济不会像存在风险市场全集的经济那样良好地运行。显然，风险市场至关重要。[16]问题是，给定风险市场的数量，经济是否能够一样的有效率？除了创造新的证券，还存在政府干预的余地吗？

戴梦德（1967）似乎提供了一个答案。如果要求企业向个人分配利润，并且分配利润是企业利润的线性函数，那么可以证明一个价值最大化的股票市场经济（在某些情形下）是帕累托有效率的。但人们不清楚的是，在一个没有完全的证券集合的世界里（从而无法为每一种不确定性和每一日期的商品定价），价值最大化的企业会如何行为。戴梦德同样为这个问题提供了答案。如果不同自然状态下（在不同的随机性下）企业的相对产出由一个特定的概率分布描述，并且企业是这个概率分布中的"价格"接受者，那么企业会相信如果将每一种自然状态下的产出增加一倍，它的价值也会增加一倍。[17]因此只需要将企业增加规模的成本与其收益相比较就可以实现价值最大化。

当然戴梦德没有回答更广泛的问题，即存在不确定性时市场的效率问题。他没有回答当企业投资更多，它的（相对）回报概率分布改变时会有什么影响。他也没有回答企业如何看待这种概率分布的改变——包括对新产品的投资。但是我的问题是，即便是在戴梦德自己假设的限制性条件下，他是否正确。在 Stiglitz（1982d）中，我证明答案是否定的。戴梦德做了一个关键的假设——似乎是一个简化的假设，我回忆起在他提交这篇文章的讨论会上，没有人反对这个假设。他假设只有一种商品。如果存在两种商品，会有什么情况发生？随着企业更多地投资，它在每一种自然状态下的产出同比例增加，

16 Borch（1962），显然，如果风险市场缺失不重要，阿罗和德布鲁就能够证明福利经济学基本定理，同时不必假设存在一个风险市场的全集。

17 戴梦德事实上利用了莫迪格利安尼和米勒（1958）的早期研究成果，后者讨论了属于不同风险等级（risk classes）的企业，一个风险等级指一个企业的回报具有特定概率分布。戴梦德假设随着企业规模改变，或者企业从事任何活动，都不改变它的风险等级；或者如果存在某种选择会影响企业的风险等级，那么会有一个明确定义的风险等级（一个明确定义的价格）对应于企业可能采取的每一种行动。我的工作证明如果没有一个完全的状态应变性证券的集合，这个假设就是不合理的。

但是产出的相对价格一般来说会改变，并且通常以非常复杂的方式改变（一般来说不是同比例改变）。一个小企业会相信它对相对价格没有影响，因此就相信本企业的市场价值会按比例增加。不难证明在这些情形下，市场配置从来都不是帕累托有效率的——即便是在戴梦德的那种限制性情形下。[18]

然而，我在更早时候就质疑过企业是否真的相信如果他们的规模增加1倍，市场价值也会增加1倍。每一个企业都认为它和其他企业有所不同，投资者也是这样认为。这就是为什么如果企业回报不完全相关，投资者会分散他们的资产组合。如果有n个相同的企业，每一个企业的股份占资产组合的1/n。如果一个企业增加规模，只有价格下降时投资者才会愿意持有额外的股份。在这种情形下，为什么企业会相信价值与规模同比例增加呢？

马科维茨（Markowitz，1952）和夏普（Sharpe，1964）的突破性工作（他们因此在1990年获得诺贝尔奖）提供了一个框架，利用这个框架可以很容易地在均值—方差世界中计算出企业的一般均衡价值。利用他们的结果很容易证明，如果不同企业的回报是独立分布的，第i个企业的价值就是

$$V_i = Xi - k\sigma_i^2$$

其中X是平均回报，而σ^2是方差。企业规模增加，比如增加1倍，会使平均回报增加1倍，但是方差会变成原来的4倍，从而企业的价值增加会小于1倍。上面的等式对于“企业投资更多时市场价值怎样变化”提供了清晰指导。在Stiglitz（1972）中，我证明市场定价（根据前面的定义）并不导致经济效率。直观的理由是简单的：保证效率的价格接受假设在这一背景下是不合理的。

上述两篇论文运用了20世纪50年代和60年代发展起来的股票市场的标准模型，探讨了在没有一个证券的完全集合条件下市场处理风险的能力。然而不是所有的风险都由股票市场来解决。人们很自然要问这个问题：市场中的个人面临风险但是没有获得完全保险，这样的市场一般来说有效率吗？比如，农民在决定农产品的种植规模时，未来农产品的产量和价格都是不确定的。给定这些不能被保险的不确定性，假设农民对于产量和价格存在理性预期，市场是有效率的吗？市场中是否存在某些政府干预——鼓励某些作物的生产而阻碍另外一些作物的生产——从而导致福利的一般性增加（帕累托改善）？戴维·纽伯利

18 人们从来都不清楚为什么“限制性”条件下的最优有这么重要的意义。比如，为什么分配利润只能是利润的线性函数？戴梦德分析的另外一个重要假设是无破产假设，我在Stiglitz（1969）和（1972a）中强调了其重要性。

（David Newbery）和我（1982）证明了的确存在这种政府干预。

这个结果同股票市场均衡效率的更早结果具有一个共同的主题：尽管小农场主（或者投资者）会合理地采用价格接受假设，事实上当他们改变投资水平时，价格的概率分布会改变。个人是风险厌恶的，由于一组完全的风险市场的缺乏，风险不能被完全消除，但是回报的概率分布的变化会影响市场转移和吸收风险的有效性。没有投资者会考虑到这一点，政府可以并且应该做到这一点，从而实现总体福利的增加。

简而言之，即便政府不能够填补风险市场缺失造成的缺口——也就是说，在市场不能为不确定性提供保险时，政府也不能提供保险——如果认识到不同部门或不同资产的投资水平会影响现存工具转移和分担风险的有效性，政府仍然可以相对于自由市场的结果改善福利。

自由贸易是合意的吗?

自从李嘉图以来，没有一个经济学概念像自由贸易那样得到最大多数经济学家的拥护。对自由贸易的支持几乎成为经济学家工会的“会员证”。每一本基础教科书都解释如何通过消除关税和其他贸易壁垒使每个国家的福利提高。尽管这种贸易自由化措施会有分配性结果，[19]但是受益者的收益会超过受损者的损失。

然而这一分析的基础是完美市场假设——包括一个风险市场的完全集合。人们很自然要问，如果不存在一个风险市场的完全集合，自由贸易还是有益的吗？这个问题不仅仅具有学术上的重要性，某些对自由贸易的反对意见正是基于如下观点：自由贸易使国家面临更大风险，而个人无法为这种风险进行保险。

纽伯利和我证明，即便在最简单的模型里，也很容易证明自由贸易会使每个人的福利下降。其中的直觉是简单的。市场本身会降低风险，这是因为当数量增加（例如，农民有好的收成），价格会下降，因此收入的波动性低于产出的波动性。在单位需求弹性下，收入不会变化。在自由贸易时，价格机制无意识地提供了完全的收入保险。但是存在自由贸易时，市场变得相对更有风险。[20]假设有两个岛屿，两个岛屿产出是完全负相关的。总产出是固

---

19 降低对相对稀缺要素的支付，同时增加对相对充裕要素的支付。参见 Samuelson - Stopler（Stopler）（1941）。

20 不仅对于消除关税的情形是这样。用关税替代配额（“关税化”）尽管经常被赞扬为 GATT/WTO 的一个伟大成就，事实上可能使发展中国家面临更大风险。参见 Dasgupta and Stiglitz（1977）。

定的。从而价格也是固定的。但是这意味着农民的收入和产出同幅度波动——收入波动性大大增加了。这转而又阻碍了对产出波动性高的部门的投资，即便平均回报很高也是这样。文章的难点是描绘出一般均衡效应，表明两个国家生产者和消费者的福利都下降了，并证明了只要两个岛屿的产出不是完全正相关，就会产生这种结果。

在发达国家（在发展中国家更是这样）风险市场远非完美，贸易自由化至少会增加某些市场参与者的风险。最起码，我和纽伯利的论文应该抑制一些对于贸易自由化的无节制的热情。[21]

几乎没有人否认我们的经济中缺乏重要的风险市场。但这显然会产生一个问题：为什么这么多风险市场都不存在？有时候，创造一个风险市场的收益小于创造风险市场的交易成本。如果是这个原因，缺少这些风险市场就不会有很大关系，福利的损失大约不会超过交易成本的规模，而交易成本（特别是在我们今天的经济中）似乎是有限的。[22]但是一般来说，个人非常在乎风险，在乎这种不安全感。缺乏保险会有很严重的社会后果——事实上，这引起对许多种社会保险的需求。

信息经济学（以及非对称信息的存在）对为什么许多风险市场缺失给出了更全面的解释。如果你比我更知道你是否患有某种隐性疾病，或者如果你故意做错事使雇主解雇你，我就不会为你的工作提供保险。因此，非对称信息的存在会破坏保险市场，从而降低福利。

三、不完美信息的福利经济学

因此，在不完美信息领域，我发起了对亚当·斯密"看不见的手"理论的最强烈批评。人们以前认识到的市场失灵尽管重要，然而范围是有限的。是的，外部性很重要，但是最重要的外部性与空气和水的污染相关，政府的环境政策借助市场机制就可以处理这些问题。不完全竞争是一个问题，但是存在很好的补救措施——反托拉斯政策，如果必要的话，政府可以对自

---

21 当然，为什么贸易自由化会降低福利还有其他解释，但它们一般被忽视了，有时候甚至被那些系统表述过它们的人所忽视。对于不完全竞争，贸易自由化可能降低福利。参见 Krugman (1980)。并且，我们指出过贸易自由化可能影响收入分配。弥补受损者所要求的收入提高会有很高代价，如果再分配的成本足够高，那么自由贸易可能不是帕累托改善的。参见，Guesnerie and Roberts (1984)。

22 这个观点不完全令人信服：我们证明由于风险市场的限制，在其他市场也会存在无效率。从某种意义上说，存在一种创造市场的外部性，市场创造所涉及的各方可能没有考虑到这一点。

然垄断实行管制。宏观经济失灵更加重要，但是货币和财政政策能够使市场经济恢复到充分就业，然后可以依赖市场机制进行处理。

与风险市场缺失相关的市场失灵更加麻烦。这种失灵更加广泛。政府可以在市场不起作用的某些重要领域提供保险——失业保险、养老金保险、犯罪保险、灾害保险。事实上，一些最热烈的关于公共政策讨论正是关于这一领域，它不仅包括养老金和失业保险改革，而且包括健康保险。尽管今天人们存在广泛共识，政府在提供一系列社会保险方面可以起到重要作用，然而目前对于政府干预应该采取的最佳形式还存在争论。[23]

但是更加重要的是信息，它是市场经济的核心。人们认为价格机制的优点是节约信息，它能够提供信号，比如，告诉生产者消费者的需求是什么。如果可以证明市场机制不能有效地处理信息，那么对市场经济的信念就会失去基础。

具有讽刺性的是，尽管长期以来，信息效率就是对市场经济进行最热烈辩护的基础，[24]阿罗和德布鲁却很少从这个角度谈论“信息”。德布鲁的一组假设与信息问题没有任何明显的联系。信息结构被视为理所当然：消费者知道产品的特征，企业知道工人的特征，贷款人知道借款人的特征。市场需要处理的唯一的信息问题是稀缺性信息。

然而，价格同时传达质量和稀缺性的信息，价格提供质量信息的功能会干扰价格提供稀缺信息的功能。如果必须提高价格来产生对企业保持质量的激励，那么这些价格就不能完全传递关于稀缺价值的信息。

有三个关键性问题：市场是否提供正确的激励使人们获取有效数量（种类）的信息？市场能否有效率地传递信息？给定信息的约束，市场是有效率的吗？后一个问题和前面关于风险市场的问题是相似的，那里我们问的是，给定风险市场的约束，市场是否有效率。

信息不是普通商品

市场经济的捍卫者，比如乔治·斯蒂格勒（George Stigler），认为信息类似于其他任何商品。[25]如果市场在提供钢材和煤炭方面是有效率的，为什

---

23 目前的争论主要是关于这些保险计划的设计，包括私人部门应该起什么作用（如果有作用的话）。例如，存在越来越多的共识，为了避免“撇脂”（cream skimming）问题（私人保险公司花费各种代价保证它们得到最好的风险，或者最定价不足（under-priced）的风险），单一支付计划（single payer program）（像美国对年满65岁的老人以及加拿大对于整个人口所使用的）可能比较好。我最近提出（Stiglitz and Yun，2002））整合各种社会保险计划有很大好处。

24 Hayek（1945）.

25 例如，参见 Stiglitz（1961）。

么在提供信息方面不会是有效率的？新的信息经济学解释了为什么斯蒂格勒是错误的：信息和其他商品有根本的不同。事实上，信息更类似于公共品，而不是私人品：增加一个人消费信息的边际成本是零；一个人拥有信息，这并不会减少其他人拥有的信息——尽管会影响其他人从这个信息中抽取租金的能力。在信息方面投资的一个关键问题是独占性（appropriability）：如果其他人可以迅速得到信息，那我就不会得到信息收益。事实上，对于可得性的考虑是格罗斯曼和我批评有效市场假设的基础[26]（参见《文集》第一卷）；如果市场传递信息是有效率的，信息可以通过价格体系在瞬间从知情者传给不知情者，那么投资于信息的人如何获得收益？如果市场是有效率的，那么唯一会被传递的信息是无成本的信息。在市场经济内，在信息和知识的使用效率与生产效率之间似乎存在一个根本性的矛盾。

可得性的问题表明对于信息经常会有投资不足。但是另一方面，对于信息也经常会有过度投资。有时候，我的信息使我以其他人为代价获得收益。在这种情形下，社会回报会低于私人回报。我比你提前一分钟知道股价将会上扬，会使我从你那里购买股票，从而得到巨额收益。但是如果提前知道此信息的是你而不是我，那么得到这些收益的就会是你而不是我。信息导致了福利的再分配，却并没有提高福利。在斯蒂格利茨（1989）中，我提供了另外一个例子：如果上课的时候，1 美元钞票掉在每一个学生的脚下，所有学生可以等到下课再去捡起这张钞票。这样是有效率的——这种拖延不会造成任何损失——但是等待并不是一个均衡。如果每一个人都在等待，我弯腰捡起我脚下的钞票就是有利的——并且可以捡起我旁边人脚下的钞票。然而每一个人都会认识到这一点，最终每一个人都会弯腰捡起钞票。唯一的均衡是每一个人都马上捡起自己的钞票，课程被打断。但是相对于课后捡起钞票，现在捡起钞票并没有改善社会福利，甚至降低了社会福利，因为课程被打断了。

在股市上面获取的许多信息都有这种性质：一个人试图在其他人之前得到信息，从而以其他人为代价获取收益。其他人当然会理性地认识到这一点，这转而会限制这种投资获取收益的程度。我不愿意与一个消息比我更灵通的人进行投机性交易。[27]结果是这些市场中的交易是有限的，从而，对信

---

26 我在一种特别的信息（由研究生产的知识）背景下充分讨论了这一点。但是在 Stiglitz（1975b）中也指出了专用性的关键问题。

27 Stiglitz（1982a）；Milgrom and Roberts（1992），更早些时候，阿克洛夫（Akerlof，1970）认为，在旧车市场中，如果市场信息不对称，交易就是单薄的或者不存在。这里可以看作他的观点的另外一个更复杂的版本。

息的投资也是有限的。在极端情形下，福利损失是有限的，因为市场中的交易和价格作用是有限的。问题是，类似的结论在真实投资（受股票交易中的价格指引）时也成立。即便信息是对称的，如果个人的偏好或环境有差别，当环境变化时也会产生交易——此时，类似的结论也成立。

存在某些条件可以解决可得性问题。如果一个信息证明我更有生产力，如果我为这个信息付费，使大家都得到这个信息，那么这就会反映在我的工资里面。显然，关于我的生产率的信息使我可以被分配到更适合我才能的工作。但是我（以及其他像我这样的人）更有才能的信息还有进一步的含义：它意味着那些没有这种信息的人更没有才能，因此他们的工资将下降。因此，像前面一样，我的某些收益是以其他人的损失为代价的。信息产生能力租金，但是由于是以其他人的损失为代价，某种含义上这意味着福利下降了——在信息导致更大不公平的时候更是这样。[28]因此，在我早期关于甄别（screening）经济学的工作中，我得出结论市场没有提供获取信息的有效激励。事实上，在最简单的模型中，可以证明存在两个均衡，一个是完全信息甄别均衡，其中每个人相对于没有甄别的均衡福利下降了。[29]

这个结果更加强健。即便开始的时候个人信息是对称的，他们仍然受到激励去获得关于他们自己的信息，这会导致信息不对称——并会导致产出和福利损失。[30]

这些并不是仅有的获取信息产生负面效应的情形。前面，我们指出过信息不对称可以破坏保险市场。如果个人不知道自己是否有容易患乳腺癌的基因，那么就会存在这种风险的保险市场。但是，如果个人可以发现自己是否有这种基因，保险公司可能不愿意与乳腺癌患者订立保险合约，因为保险公司明白：那些最愿意购买这种保险的人是那些知道他们有这种基因的人。如果要依靠私人提供这种保险，那么必须禁止个人获得这种信息。[31]

信息披露的激励

当然，市场的确提供了一些披露信息的激励。好的产品总想得到更高的价格。好的工人想得到与之相配的更高的工资。如果披露可以被证实并且没有成本，在这样的世界中，唯一的均衡是存在这种披露。我们很容易证明这

28　假定存在一个平等主义的社会福利函数。

29　Stiglitz（1975b）.

30　Stiglitz（1984）.

31　参见 Rothschild and Stiglitz（1982，1997）。

一点。当然，最没有生产力的个人或者最低质量的生产者没有动力披露信息，但是在我称为瓦尔拉斯甄别法则（Walvas of Law Screening）的过程中，如果除了一个人以外的每一个人都披露了自身能力（质量）信息，那么在理性预期下，最后一个人的能力（质量）也被揭示了。信息会被完全揭示。但是在真实世界中，披露信息是有代价的，证实被披露的信息也是有代价的。[32]在这个世界中，显示/披露的信息数量可能不是最优的。政府也可以起一定作用，例如，可以强迫当事人披露信息（特别是在市场非竞争性的情形下）。

尽管我关心的是市场提供有效数量信息的激励，我更关心当信息有限时市场是否会做出生产多少的正确决定，以及产生的信息和市场经济价格之间的相互作用。在我对风险市场的分析中，我指出过生产决定影响了市场传递和吸收风险的能力，但是没有生产者将这种影响考虑在内。这就是为什么风险市场数量有限的经济是没有效率的。我猜测这里也存在类似的现象。在一系列论文中，我们证明事实上就是这样。

在基本的自选择模型中可以看到我对这个问题的基本见解。在《文集》第一卷，我阐述了可以将表面相似的个人区分开来的两种方式。一种是让他们参加考试（我正式地称之为甄别），另一种是让他们面临选择，通过他们所做的选择，显示自己是什么样的人。自选择产生所需要的信息：更可能发生事故的个人选择更小免赔额的保险单，更有能力的个人选择更多按绩效付报酬的劳动合约，好企业的所有者更多地依赖借债筹集资本，能力较强的学生待在学校时间更长，因为这样做的成本更低。自选择也有约束。例如，能力较低的学生在低收入/低教育水平之间无差异，能力较强的学生获得更高的收入和更高的教育水平。自选择约束“促使”更有能力的学生得到他们希望的更高的教育水平。市场似乎是帕累托有效率的——也就是说，不让他们经受这种扭曲（或者代价更高的甄别过程）就无法识别出那些更有能力的学生。但是更进一步的思考表明：还有许多其他方式，能够用更低的成本达到甄别的目的。继续我们的故事，假设在学校里，聪明的学生用头脑进行所有的计算，而不聪明的学生必须用铅笔。增加铅笔的成本会使自选择约束更有“约束力”（binding），这反而增加了教育体系作为甄别/信号系统的效率。税收会产生小的（二阶的）扭曲（与税收的大小成正比），但是高能力学生可以减少“自选择”所需要的教

---

32 当存在某些市场权力容易受到信息披露不利影响时，情况更是如此。

育数量，因此减少了一阶的扭曲。任何干预，只要对两个群体的无差异曲线产生不同影响，就能导致帕累托改进。这背后的简单数学推导可以参见 Arnott，Greenwald and Stiglitz（1995）。

同样的见解对于存在激励和道德风险问题的经济也成立。为了避免个人承担过量风险（完全保险的时候就会这样，因为被保险人不承担风险成本），市场提供了有限的保险。看起来市场能够对更多转移风险的收益和减少激励的损失进行平衡，从而市场均衡会是有效率的，事实上，在一些高度限制性的条件下，我们可以证明这一点。[33]

与分析上述选择的自选择约束相对应，理查德·阿诺特和我发展了一个分析性框架。[34]假设一个人有两组无差异曲线，一组描述他抽烟时的偏好，另外一组描述不抽烟时的偏好。在零利润约束和个人不会从不抽烟“转向”抽烟的约束下（这里假设了抽烟增加被保险事件发生的风险，比如癌症），最优保险合约提供了最高水平的保险。在这个限制性的框架内，政府没有其他选择，市场变得有效率。但是政府还有其他工具。政府可以向香烟或者其他抽烟的互补品征税，并为抽烟的替代品提供补贴。对商品市场的这种干预可以使风险市场提供更多的保险，同时也不让个人转向抽烟。显然，保险市场上发生的事件无法同产品市场上发生的事件完全分离。

基于同样理由，为一种风险提供的保险数量可能影响个人承担另一种风险的意愿。阿罗和德布鲁对市场经济的分析似乎表明很高程度的分散化是可能的。是的，所有的市场都是相互联系的，当每一行业的企业最大化它们自己的利润时，就会带来经济效率（或者更精确地说，经济是帕累托有效率的）。但是当信息不完美时，分散化似乎不能很好地起作用。在一种商品或者保险市场上发生的事件对于其他保险市场有一阶的影响。同样，直觉是简单的。当其他市场有效率时，价格等于边际成本，或者商品的边际社会收益等于边际社会成本。对于商品需求的变化会导致生产略微增加，其福利效应可以忽略，这是因为在边际上，边际社会收益等于边际社会成本。但是如果消费者受保险数量的约束（这是因为保险公司担心提供更多保险，所有的高风险个人会购买风险），那么在消费者从事风险性活动以前，如果一个边际变化增加了他们所需求的保险数量——从而对于高风险个人有二阶的影响——就会改变提供给低风险个人的保险数量，从而导致他们的福利一阶

33 Shavell（1979）.

34 Arnott and Stiglitz（1988）.

改善。

布鲁斯·格林瓦尔德和我在20世纪80年代中期发展了一个分析框架，探讨这些问题以及其他许多与不完美信息信息成本、不完全市场和交易成本有关的市场不完美问题。在某种意义上，它包含了之前和之后的许多结果。我们的想法是在这些市场中存在某种金钱外部性（pecuniary externality）。传统的马歇尔早期分析区分了技术外部性和金钱外部性。比如，邻居家的苹果园影响蜜蜂产蜜的能力。这是真实的或者技术的外部性。另一方面，如果企业更多生产某种商品，它可能压低价格，这会影响其他企业，这种影响叫做金钱外部性。在运行良好的竞争性市场经济中，技术外部性很重要，但是金钱外部性不重要，因为在均衡时，一切都是最优的。我们的工作证明这些金钱外部性存在真实影响。如果香烟价格提高导致抽烟下降，癌症减少，那么保险费会降低。由香烟消费下降带来的净损失（dead weight loss）小于保费下降带来的收益。税收降低了以前存在的扭曲的程度。个人决定是否抽烟时只考虑他们需要承担的成本，而不考虑保险公司承担的成本。这是很大的扭曲，我们有必要降低这种扭曲，即便是以增加另外一种扭曲为代价（通过征取少量的税收）。[35]

我们证明如何用同一个框架探讨道德风险和逆向选择问题，以及存在缺失市场时的问题。在 Greenwald and Stiglitz（1988）中，我们将此框架扩展到搜寻和效率工资模型。

格林瓦尔德和斯蒂格利茨的研究代表了福利经济学中一场静悄悄的革命。在此之前，人们认为市场会导致有效率的资源配置。市场的伟大优点是能够分散化——一个部门的人不需要注意其他部门的情况。所有相关的信息通过价格体系来传递。正如我们指出的，存在某些情形——当存在污染、公共品、或者竞争不完全的时候——市场运行并不良好。信息效率似乎处于市场经济成功的核心地位，但是我们描述的信息不完美是普遍的。对于前面提到的市场失灵，可以通过一些简单的政府干预加以校正；但是由于信息不完美在经济中普遍存在，需要政府更加广泛地干预市场。

正如我前面指出的，阿罗和德布鲁发现的市场有效率的条件是一个奇异集合。他们强调有限的一组信息问题——而忽略了关键的信息问题，比如工人质量、产品、企业、观念……在这个更广泛的领域，有人认为市场是无效

35 显然，这个结果同拉姆齐（Ramsey，1928）对最优税收的分析类似。拉姆齐证明很多小的扭曲优于一个大的扭曲，因此，不能够只靠计算扭曲数量评估一个市场的效率。

率的，分散化（至少像长久以来被认为那样）是不理想的。[36]政府可以起到重要的作用。但政府是否能够做到则是另外一个问题。当市场有效率时，经济学家不需要过多注意政府。不管政府运转多么良好，没有一个政府可以改进市场的运行。现在很清楚，市场失灵了，而且非常普遍。我们需要更多了解政府的行为，理解政府可以在何种程度上改善福利，但是我们应该清楚的是，某些增加福利的政府干预是很有限的，比如对香烟征税。[37]

在一些更通俗的著作里，[38]我提请人们注意这些市场失灵在20世纪90年代美国（以及其他国家）面对的经济问题中起到了重要作用。经理（CEO）使用管理权力，引入管理激励计划和一些会计制度，这些计划和制度促使他们提供扭曲的信息，使股东很难分辨经理行为的真正结果——从而无法对股票合理定价。会计公司代表持股人监督企业的激励下降了，相反它们同企业经理合作，一起提供扭曲的信息。投资银行在提供信息和其他行为（比如IPO的价格）方面同样受到的扭曲激励。尽管这里的模型看起来很抽象，它们对于真实世界具有重要的含义。某些干预会以很低的成本带来巨大的收益。披露要求和对企业行为的某些约束可以增加经济的信息效率，降低利益冲突的程度。这方面的例子包括：对会计公司加以约束，强迫它们集中于会计而不是咨询，以及对于股票期权的披露要求。

非市场的制度可以驯服和调节市场吗？

随着对政府信心的下降，许多经济学家认为即便市场失灵，人们也会创造非市场的制度来填补空缺。在很多领域，我们看到似乎的确是这样。我们看到信息不对称如何导致有限的保险市场。由于市场只提供一部分保险，个人转向其他领域寻求风险分担。比如，家庭在风险分担中起到重要作用。如果我得病并且不再有收入，我妻子会继续为我提供食品和住宿（至少在一

36　在其他应用性更强的研究中，我证明了激励效应如何限制分散化的程度。布雷弗曼（Braverman）和我分析了农村市场，指出通常信贷、劳动和产品市场是相互联系的。我们证明了如何用这种联系“控制”道德风险问题。参见 Braverman and Stiglitz（1982），在 Braverman and Stiglitz（1986）中我们做了进一步扩展，包括了肥料和其他投入。Andy Weiss and Stiglitz（1983）进一步拓展到跨期联系的情形。这种联系限制了跨期竞争的程度。签订了长期合约的企业更缺乏自由寻找更能满足它们要求的替代品。

37　从而格林瓦尔德和我证明，当考虑进信息成本和市场中存在的约束，市场不是受约束的帕累托有效。显然，如果信息是完美的（获得信息没有成本时有可能是这样），有可能会增加福利。但这不是我们提出过的问题。

38　特别请参见《喧嚣的九十年代》（*Roaring Nineties*，2003b，中文版由中国金融出版社出版）。

段时间内)。在早期，产生了许多提供保险的合作性活动，比如合作性丧葬社团（burial societies)。许多保险公司的前身都是相互保险社团。

非市场制度填补市场空缺并执行市场的某些功能并不意味着它们真的起到了作用，至少不一定很好地起作用（这有时候称作功能谬误（functional fallacy))。非市场制度产生的时候，市场制度会做出反应。这里的问题是，这种非市场制度是否能够增加福利？阿诺特和我（1991a）得出了一个很站得住脚的结论结果（尽管很大程度上是一个“二手经济学家”回答)。从另一方面来说，如果家庭成员不监督其他成员的行为，保险公司会发现在市场提供的任何一种保险水平上，它会承担更多的风险，因而保险公司会降低风险覆盖范围，提高保险费，其净效应是非市场保险将市场保险挤出，使每一个人的福利下降。尽管非市场制度的动机是良好的，但会有功能障碍（dysfunction)。如果家庭成员可以很紧密地互相监督（保证他们可以晚上睡眠充足，在下雨时候不外出，等等)，那么结果会有显著不同。这是因为家庭成员不得不为“坏的行为”承担某些成本，他们有激励进行这种监督，并且至少在某些家庭，他们有能力这样做。在这些情形下，福利增加了，坏的行为被减少了，对于被保险人的行为改善来说，保险公司本质上是一个搭便车者。

阿诺特和我的上述文章分析了一种重要的相互作用——市场和非市场制度之间的相互作用——遗憾的是到目前为止这个主题很少得到经济学家的注意。正如非市场制度可能损害市场，市场制度也可能损害非市场制度，比如许多发展中国家中存在的管制公共资源的社会控制机制。

四、创新

市场的支持者夸耀市场的信息效率，然而过去30年间的研究证明市场的许多最严重限制来源于不完美和不对称信息问题。格林瓦尔德和斯蒂格利茨（1986）表明每当存在这些信息不完美——几乎总是这样——市场甚至不是受约束的帕累托有效率。

市场支持者还经常夸耀市场在创新方面的成功。但是创新导致了知识（一种特殊的信息）的生产。正如我在20世纪70年代早期指出的，与信息的生产使用有关的大多数市场失灵问题也同与创新相关的知识的生产使用有关。[39]当然，即便是赞成市场的人，也认为需要政府——以保护知识产权。

39 特别参见Stiglitz（1975)，在Stiglitz（1987）中，我讨论了作为公共品的知识以及可得性的问题。

但是知识产权限制了知识的使用，创造了垄断权力——二者都产生市场无效率，这些静态的无效率必须和研发激励增加产生的动态收益相平衡。

知识，特别是基础研究，是一种公共品，因此很多人强烈支持对于基础研究的公共支持。应该将知识如何融资和知识如何生产两个问题区分开来。政府应该为大学和研究机构（包括私人和公共的）里面的研究提供资金。然而这个问题是一个经典的委托—代理问题：人们如何保证那些被资助者努力工作？可能很难对其进行监督，并且由于结果有很大不确定性，用投入来判断产出可能是不合适的。然而，政府和大学管理实验室的制度在许多基础研究领域都取得了广泛成功。激励科学家的是对知识的渴望，以及做出重要发现带来的赞誉。货币报酬只是激励结构的一部分。

某些经济学家强调知识产权在为研究提供激励中的作用。最近的文献表明，因为社会和私人回报有显著的不同，即便是设计良好的知识产权制度也会导致激励的扭曲。企业先于其他所有人发现一个构思（idea）的社会回报只是那项构思（产品）相对于以前构思增加的边际价值。这里存在过多发放专利的风险。任何专利权都提高了利用知识的成本，而知识是生产更多知识的最重要投入。提高知识的成本可能降低创新的步伐。许多专利（至少是部分）包含有现存的知识。在这种情形下，专利代表“公共领域的圈地运动”，攫取公共空间的东西，将其放入私人空间。尽管有人认为公共土地的私有化会增加效率，在我们这个例子中，对公共品（知识）的私有化显然会降低效率。在我最近一本书（Making Globalization Work）[40]中，我提出了一系列其他原因，说明为什么专利制度能够并且已经减慢了创新的步伐。专利丛林（Patent thickets）——相互冲突的专利要求权——使得创新不再有吸引力，因为任何创新者都面临支付巨额法律费用的风险。垄断会减少竞争的激励，而专利有助于创造垄断。许多人认为微软的行为模式——在操作系统领域使用市场权力剽窃并/或压制创新，像浏览器和媒体播放器那样，特别是当这些创新可能损害其市场权力时——对于软件业许多领域的创新都有窒息作用。

对目前的知识产权制度的另外一种选择是奖励制度——至少在那些研究目标明确的领域是这样，比如发明疫苗或者治疗疟疾。那些做出发现的人会得到奖励，然后所有人都可以使用这种知识（也许需要支付使用费）。这种制度利用了竞争市场，保证发明的好处得以广泛传播——比如以低成本提供

40 Stiglitz（W. W. Norton，2006）.

药物；相反，在目前的专利制度下，垄断者有激励提高价格并且限制知识带来的收益。

在20世纪的前半叶，熊彼特（Schumpeter）指出对创新的激励是市场的一个伟大优点。他认为存在一系列的暂时性垄断者，对成为下一个垄断者的竞争提供了创新的关键性动力。然而对激励的更进一步研究表明：垄断者有动力去购买竞争者的研究，从而扩展它们的垄断权力，垄断权力远非暂时性的。产权制度经常成为企业扩展垄断权力策略的一部分。潜在竞争的力量可能是有限的：垄断者受到激励是远远超过竞争者，从而使竞争者很难蛙跳（leapfrog），垄断者一旦确立领导者地位，他们就可以放慢研究和创新的步伐。

从对创新的分析中得到了一些教训：第一个是在资助研究方面可以起到重要作用，不应该过分依赖知识产权。需要一个政策组合——包括由政府支持的基础研究、奖励制度和标准的知识产权制度。拥有知识产权不代表得到许可滥用这种权利；在某些情形下，政府可以发放强制性许可证（compulsory license）。

五、政府的职能

当市场失灵时，很自然会问，政府可以做什么改进？政府的职能是什么？作为一个学者和政策制定者，我大部分的职业生涯被这个问题所吸引。这里提到的研究论文试图说明一些一般原理——更多是关于政府应该做什么，而不是它事实上做了什么。因此，这些文章是属于福利经济学的传统范围。

在非常广泛的水平上，市场失灵方法提供了一个思考政府职能的简单框架：政府应该干预市场，消除或减少市场失灵的程度。用这里提出的模型来看，政府应该征税和补贴，达到受约束的帕累托改进。

例如，在 Stiglitz（1993d）中，我指出了金融市场中的许多与信息相关的问题。对于银行资产组合头寸的信息——违约风险——是至关重要的公共品。让每一个存款者仔细研读资产负债表或者雇用其他人这样做几乎没有意义（私人信用评级机构可能提供类似或者补充的服务，但是提供这些服务的市场远非竞争性的——几个机构主宰了市场——并且它们的历史表现并不好，比如在最近的东亚危机中；[41] 它们使审查对象进行合作——完全和诚实

41 "The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis", G. Ferri, Liu and Stiglitz, *Economic Notes*, 28 (3), 1999, pp. 335–355.

的披露——的工具也许不够，此外它们还可能会面临利益冲突。例如，某些机构提供服务帮助它们的审查对象改善信用评级）。因此，认为存款保险减弱了存款人监督银行的激励从而应该取消存款保险[42]的观点是完全错误的。当然应该小心地检查那些资产迅速增加的银行，特别是当它们支付的利率远远高于普通市场利率时。它们发现了其他贷款人没有发现的新的高收益投资，还是它们改善了交易效率从而在不损害生存能力的情况下能够支付更高的利率？或者是他们以公共为代价承担了不适当的风险，知道如果他们失败，会有政府援助？银行管制者应该限制银行可以支付的利率，相信这种高利率与银行承担更高的风险存在系统相关。[43]

在 Stiglitz（1993b）中，我探讨了更广泛的问题：为什么政府今天在金融和非金融部门的风险承担中起到这么大的作用，我指出，部分是因为监督是一种公共品，部分是因为如果监督是政府责任，政府不可避免地要承担错误的后果。但是上述文章提请人们注意至少两个其他因素：第一，许多风险是社会风险，社会中许多人都同时面对这种风险。这种情况下，私人保险就是不可行的。事实上，通过政府，这一代人可以向未来的人借款。因此，许多人想购买通货膨胀保险，但是他们向谁购买保险呢？如果通货膨胀上升，数亿万美元会用于赔付。由社会中许多人同时面对的风险有时候称为社会风险。第二，存在承诺（commitment）的问题。个人知道如果他们不购买保险，政府仍然会进行援助。而社会发现某种水平的伤害是不可忍受的——不管原因是什么，即便这种伤害是自我导致（self-inflicted）的。由于这个原因，会存在道德风险问题。一种仁慈的处理道德风险的方式是要求个人购买足量保险，从而使政府援助变得不必要。

政府干预的设计

尽管保守主义意识形态继续认为市场本身可以解决所有的经济问题，对市场失灵的分析已经表明市场的许多局限。在某些部门，比如健康保险，这些局限性非常显著。在前面描述的许多例子中，市场根本无法提供个人需要的产品（失业保险）。

---

42 国际金融机构经常怂恿发达国家采取这种立场。

43 我创立了实行银行管制的一种更综合的方法，请特别参见“Robust Financial Restraint”，in *Financial Liberalization*：*How Far*，*How Fast*? G. Caprio，P. Honohan and J. Stiglitz 2001a and 2001b（eds.），Cambridge，UK：Cambridge University Press，2001，pp. 31－63；“Principles of Financial Regulation：A：Dynamic Approach”，*The World Bank Observer*，16（1），Spring 2001，pp. 1-18。

今天的争论（至少是学者之间）不是市场是否失灵，而是政府能否纠正这些市场失灵，或者至少增加福利。当然，许多政府项目存在无效率，某些受腐败的困扰——然而政府腐败的规模几乎从未达到过20世纪90年代震动美国的金融丑闻中展示的那种私人部门腐败的规模。许多政府项目有很高的效率。相对于私人企业提供的年金保险，美国的社会保险计划更加有效率（有更低的交易成本），提供了私人提供者无法提供的风险（比如通胀）保险。韩国政府的钢铁公司比美国的大多数（如果不是所有）私人钢铁公司有效率的多。

在几乎每一个成功的经济中，政府都起到了巨大的作用。比如在美国，政府在金融（小企业贷款、住房抵押贷款、学生贷款）和研究中起了巨大作用。互联网，20世纪后半叶最重要的创新，很大程度上是由美国政府构想和资助的。

今天的研究和政策的一个主要目标是提高公共部门的效率，并且减少"政府失灵"的范围。例如，政府已经学会如何使用拍卖更有效地销售资产或者进行采购，它们已经学会如何更广泛地使用激励结构。

早期对政府职能的讨论集中于哪些部门应该由政府负责，哪些部门应该留给私人来做。今天，这个问题变得完全不同：政府和私人部门经常参与相同的部门，各自都有各自的作用——它们可以被认为是互补的。例如，在金融部门中，政府在管制方面有关键性的作用；在许多国家，政府介入以填补"缺失的市场"，例如，对学生和小企业的贷款。在研究领域，政府对基础研究有特别重要的作用，而私人企业在应用性研究方面扮演更重要的角色。政府和私人部门的最佳组合在国与国之间会有不同，这取决于公共部门和私人部门各自的优缺点。

最近人们思想的另外一个重要转变是认识到"第三部门"（非政府、非盈利制度）的重要性，比如，大学、基金、合作社和市民社团。它们在某些领域，特别是教育、健康和研究中起了关键作用。

得到广泛注意的另外一个问题是政府干预的层面——局部的、全国的、地区的或者全球的。对于局部、国家和全球公共品的分析——在《文集》第五卷中讨论过——我提供了重要的见解：某些产品的利益集中在某一个社区内，这时候应该由地区做出提供这种产品的决定。其他一些公共品的好处是全球性的。对于全球化的部分批评是过于经常地执行全球标准（例如，在知识产权领域），而事实上在不同环境、不同国家中实行不同标准有显著的优点：统一标准的成本可能超过收益。

再分配问题可以与效率问题分开吗？

即便市场是帕累托有效率的，它们的福利分配也可能很不公平，以至于社会无法接受。那些没有技能或资本的人会处于生存水平以下。因此，在再分配中，政府发挥着重要作用。

现代福利经济学——事实上是所有现代经济学——的一个主要问题是关于区分公平和效率问题的能力。经济学家喜欢强调效率问题——市场、对利润的追求以及价格制度导致有效率的结果。“公平”问题可以留给政治过程。任何理想的福利分配可以通过市场得到；政府需要做的只是进行一次性再分配以达到理想的分配水平。正如面对市场失灵时，需要有限的干预以得到经济效率，也需要有限的干预以得到“社会公正”。

在政治层面，我将这种公平—效率二分法看作忽视不平等问题的一种借口。当经济学家赞成一种导致贫困或不平等程度恶化的政策时，这个借口使他们可以心安理得。问题不是来自他们给出的建议，而是来自政治过程，显然他们不能为政治过程负责任。

但是更深一层，随着信息经济学提供了对于经济过程的更深刻理解，我更加相信这种新古典二分法在经济上是不成立的。

我们对这些问题的理解有几个主要突破。第一个出现在莫里斯（Mirrlees，1971）的著作中，以及我在《文集》第五卷的研究中。[44]一般来说，新古典二分法设想的无成本的一次性再分配并不可行，因为政府不具备必要的信息——基于这种信息它可以决定谁应该进行一次性支付，谁应该接受这种一次性支付。由此，我们迈出了破坏公平和效率完全分离的第一小步。[45]如果政府缺乏这种信息，那么政府不得不将净转移基于可观测变量，比如收入，但是这样做会不可避免地导致扭曲。而这种扭曲程度取决于市场造成的不平等程度。因此，干预市场以改变市场的收入分配很可能是有益的，这样

44 莫里斯强调最优税收的概念，这种税收结构最大化了（功利主义）社会福利函数。我发展了更一般的概念——帕累托有效率的税收结构，在这种税收结构下，在不使任何人福利下降情况下无法使任何人福利提高。从某种意义上说，我回到了更一般的古典问题——什么是有效率的税收结构。参见 Stiglitz（1982c，中文译文见《文集》第五卷）；Stiglitz（1987c）；Brito，Hamilton，Slutsky and Stiglitz（1990）.

45 特别请参见“Pareto Efficient Taxation and Expenditure Policies，With Applications to the Taxation of Capital，Public Investment，and Externalities，”1998 年 1 月在向桑德莫（Agnar Sandmo）致意的讨论会上提交。莫里斯和随后的研究假设熟练工人和不熟练工人之间的相对工资是固定的，因此在他的模型中，无法探讨事前的分配政策和事后的再分配政策之间的权衡。在我的论文中，收入分配是内生的，受政府政策影响。

做可以降低事前的不平等程度，从而减少对扭曲性的政府再分配的压力。

但是收入分配可以通过更深刻的方式影响经济效率：我早期工作关心的一个核心问题是代理关系。[46]在分成制度下佃农的激励减少了。给定不平等的程度，分成制度可能是有效率的经济关系[47]；但是，工人只能得到他们产出的一半，这的确减弱了激励。产生代理关系是由于在物质财富（这里是土地）所有权和“劳动力”所有权之间的分离。土地再分配会消除对代理关系的需求，同时消除它所产生的问题。例如，缺少财富会阻碍工人提供担保（post bond）的能力，从而加剧工业背景下的偷懒问题。[48]另一方面，存在某些报酬递增的情形，需要集中大量财富。当财富所有权更加分散时，这也会产生一种在财富更加集中时不会存在的代理关系。[49]

比较经济制度

福利经济学可以被看作是更广泛的比较经济制度和组织设计研究的一部分。它试图评估一种特定的经济制度——竞争市场制度。

市场经济的一个关键特征是分散化。分散化由于节约信息而受到赞扬。企业只需要知道它们自己的技术，它们自己工人的特征等。所有其他相关的信息可以通过价格传递。企业不需要知道数量（市场上的总数量，或者竞争者的生产决定）；也不需要知道消费者偏好的任何信息。然而，新的信息范式破坏了依赖分散化的能力。如前面所引用过的，格林瓦尔德和斯蒂格利茨（1986）[50]的一个结论是存在某些重要情形，在其中无法通过分散化得到有效率的结果。类似地，阿诺特和我证明以分散的方式有效率地提供保险是不可能的。在1991年同阿诺特合作的一篇未发表的论文中，我证明由于承担风险的数量与总的风险覆盖有关，一个分散化的市场——其中保险由许多人分别提供，并且因此对于加总的数量信息不可得——不会是有效率的。在其他存在道德风险（激励）问题

46 “Incentives and Risk Sharing in Sharecropping”, *Review of Economic Studies*, 41. April 1974, pp. 219 – 255.

47 但是正如我在这一节提到的论文所表明的，市场均衡仍然不是完全有效率的，即甚至不是受约束的帕累托有效率。

48 Shaprio and Stiglitz（1984）.

49 对这些问题的进一步讨论，参见Hoff（1996）。这一节的第一组论文与福利经济学第一基本定理有关，认为每一个竞争均衡都是帕累托有效率的。刚才描述的工作更多与第二基本定理相关，它断言给定初始的收入分配，每一个帕累托有效产出都可以通过市场过程达到。第二定理所要求的主要数学性质是“凸性”。我们对信息经济学的早期工作表明，在信息内生化的经济中，非凸性非常普遍。可参见我的诺贝尔获奖演讲；Arnott and Stiglitz（1988）；Radner and Stiglitz（1984）。

50 也请参见前面引用过的Braverman and Stiglitz（1982）。

的市场环境中（包括金融市场），也会产生类似的问题。

关于信息和经济组织的《文集》的第二卷中的论文强调了分散化的另外一个方面：独立决策者的优点。所有的决策者都可能犯错误。结果是好的项目会被拒绝，坏的项目会被接受。当好项目缺乏之时——因此人们更担心好的项目被拒绝，而不是坏项目被接受——分散化特别重要。[51]

比较经济制度还强调将私有产权的市场经济和市场社会主义作对比。对价格制度的早期分析[52]表明在市场经济和使用市场机制的社会主义经济之间很少或没有差别。在两种情形下，价格都会有效率地指导资源配置。在实践中，市场社会主义从来都无法顺利运行。我在《市场社会主义和新古典经济学》（"*Market Socialism and Neoclassical Economics*"）[53,54]中指出，真正的问题在于所使用的市场经济模型。竞争经济的阿罗—德布鲁模型忽视了中心问题——激励问题，基本的信息问题（而不是对稀缺资源的利用），包括那些与质量和创新发明相关的信息问题。如果阿罗—德布鲁模型是正确的，如果它正确描述了市场经济的运行，则市场社会主义会更加成功。

许多人认为市场社会主义的失败关键是缺乏私人产权。因此他们认为，私有化应该处于从共产主义向市场经济转型的核心位置。这提出了根本性的问题：什么时候私有化会明显改善福利。我关注的不是对钢或者其他存在竞争性市场的传统"私人"品的私有化；私有化的鼓吹者远远超出了这一点，私有化的范围包括自然垄断和社会服务，从水到有毒物质。近年来对这种私有化的经验是混合的。大卫·夏平顿（David Sappington）和我（1987）提出了更深层次的问题，[55]在哪些条件下我们预期私有化会成功。我们的分析对于那些推动私有化的人是一个重要的打击。结果表明私有化所需条件是高度限制性的，与那些证明福利经济学基本定理（保证市场经济效率）所需要的条件相类似。他们这样做表明了，意识形态和经济科学一样在推动私有化方面起了重要的作用。[56]

---

51 特别请参见 Sah and Stiglitz（1985，Stiglitz 1986）。

52 Oscar Lange（1964）.

53 Stiglitz（1993a）编辑过的版本。

54 也请参见 *Whither Socialism*? Cambridge，MA：MIT Press，1994（这本书是由 1990 年 5 月维克塞尔讲座的一篇论文扩展而来的）。

55 Sappington and Stiglitz（1987），中文译文见《斯蒂格利茨经济学文集》第二卷。

56 私有化方面更多应用性的问题和向市场经济的转型在 *Whither Socialism*?（1994）中有讨论，另参见 Stiglitz（1993c，1993e）。

最后，从共产主义向市场经济的过渡为比较经济制度/战略提供了一个最重要的“真实生活”的案例。中国和俄国采取了完全不同的战略，后者的战略导致了比共产主义下更坏的绩效，前者导致了显著更好的绩效。解释这种经验，并从中提取它对于政策和理论的经验教训，显然是未来面对经济学家的主要挑战之一。[57]这些经验包括：（a）保证竞争比私有化更加重要；(b）保证充分就业可能比保证低通货膨胀更加重要；（c）创造新企业可能比重组老企业更加重要；（d）不完美和实用的解决方法——比如个人责任制度，不会带来完全的私有化——可能比意识形态驱动的解决方案更加有效；（e）自由化和私有化是达到目的的手段，而不是目的本身；在这些改革的速度和顺序方面必须小心；（f）公司治理很重要；（g）设计拙劣的改革（比如不合法的私有化）可能破坏对改革的支持和法律规则的创造。

六、未来研究的方向

显然格林·瓦尔德—斯蒂格利茨的论文结论是强有力并具有一般性的，但在很多团体中仍然持续存在支持市场效率的观点，这是很奇怪的。许多人试图提出其他理论模型，在那些理论模型中即便有信息不对称，市场仍然是(受约束的）帕累托有效率。[58]但是以我的观点，以及经济学家的普遍观点，这些努力并没有取得成功。

由此可以得出结论，市场不是受约束的帕累托有效率。这产生了四种可能性。第一种是尽管无效率非常普遍，但是程度很低。我对此表示怀疑，但这会是未来一个重要且合理的研究方向。我相信无效率的影响显著的原因很简单，只用一个例子就可以解释：我们前面阐述过不对称信息会导致不完美的风险市场，例如，缺乏保险以及有限的股权市场。[59]在有限的股权市场下，企业的行为方式会表现出非常的厌恶风险。[60]这种对风险承担的限制对于经

57 关于人们早期的努力，参见 Stiglitz（2000a）；Stiglitz（2000b）；Hussain，Stern 和 and Stiglitz（2000）；Ellerman and Stiglitz（2000），(2001）。

58 Prescott（1984）.

59 Bruce Greenwald，J. E. Stiglitz，and A. Weiss，“Informational Imperfections in the Capital Markets and Macroeconomic Fluctuations，” *American Economic Review*，74（2），May 1984，pp. 194 – 199；J. E. Stiglitz，“Information and Capital Markets，” In *Financial Economics*：*Essays in Honor of Paul Cootner*，William F. Sharpe and Cathryn Cootner（eds.），Prentice Hall，New Jersey，1982，pp. 118 – 158i；Meyers and Maljuf (1996)；S. Ross（1973）；and Leland and Pyle（1977）.

60 Greenwald and Stiglitz（1990）；Stiglitz（1989c）.

济演化有根本的影响。

今天，美国卫生保健部门支出超过了 GDP 的 14%，这方面支出的大部分是通过公共或者私人保险提供。保险是很不完全的，美国人口的很大部分没有保险。人们有这样一个共识，保险和缺乏保险一样都有巨额的成本，原因在于保险导致对于卫生保健的过量支出。

第二种可能是我和阿诺特的工作低估了非市场制度填补市场缺口的潜力。在一些情形中，非市场制度的确起到了作用。在《同业监督与信用市场》（“Peer Monitoring and Credit Markets”，1990，中文译文见《文集》第二卷）中，我描述了 Grameen 银行，一个非营利的微观信贷机构，利用同业监督实现了很高的偿还率。这克服了以前困扰农村信贷市场的一些不完美信息问题。这一领域的进一步研究可以探索一些其他方式——通过这些方式，非市场制度可以有助于改进市场配置的效率。但是正如我前面指出过的，我也关心市场制度对于非市场机制配置资源的负面影响。同样，这两者都代表了潜在的很重要的研究领域。

第三种可能是尽管市场会失灵，但政府失灵更加严重。我在前面承认，我的确看到了政府失灵。但我还看到了政府成功，在其他地方[61]我不但描述了政府的成功，而且描述了通过什么方式可以使政府更有效和更有效率。例如，不完美信息在私人部门是一个问题，在公共部门同样是一个问题。然而，规制和处理可以影响信息披露，可以限制与信息相关的政府失灵；而更高的透明性会带来更有效的公共控制。[62]更广泛地，我们需要更多和更好的对公共部门的研究，研究政府部门克服市场失灵的能力的局限性。[63]

最后，我们将我们的分析放在一个在经济学中占 100 多年主流地位的经济分析框架内——瓦尔拉斯模型，该模型存在假设理性的消费者和利润最大化的企业（或者至少是效用最大化的经理）。许多自然的扩展（不理性的消费者、管理人控制的企业）可能只会加强市场局限性的观点。但是另外一类的文献——演化经济学——其结论对某些人来说显得更加乐观。

由于经济学家认识到标准均衡模型的限制，他们开始寻找其他模型，包括演化模型。某些作者似乎认为演化具有某些“最优”性质。当然这是社

61 Stiglitz（1991）；Stiglitz（1997）.

62 Stiglitz（2003a）；J. E. Stiglitz，“Information and Political Processes”（即将发表）。

63 例如，政府行为的某些流行的模型研究寻租行为时，假设寻租市场是完全竞争的，从而租金都被消散掉了。但是像经济市场一样，政治市场远不是完全可竞争的。在有限竞争下，可能有最小的租金耗散，与政府保护相关的总福利损失可能是哈伯格（Harberger）三角，且通常是有限的。

会达尔文主义者的思想，他们认为自然选择清除了那些弱者和无能力者。在许多领域，这种达尔文精神潜伏在残酷的竞争信念的背后。尽管这些概念对某些人来说是有吸引力的（并且对其他人来说是可恶的），至少在经济学范围内，并没有理论模型证明演化过程的效率（在任何广为接受的意义上）。我确信，在瓦尔拉斯框架内限制市场过程效率的同样力量，也会限制演化背景下的效率。例如，考虑资本市场的不完美性。一个有长期生存价值（survival value）的物种不能靠未来的繁荣借款。如果某个物种能更好地适应今天的环境，它可能排挤出更适应长期生存的物种，使其灭绝。一个物种的存活依赖于生存的其他物种的种类。存活是“一般均衡”问题的一部分，但是演化过程经常在“部分均衡”背景下运行，也就是说，在一定环境下，两个物种中哪一个表现更好。结果可能会有多重均衡，其中一个物种（至少是在特定含义上）好于其他物种。在一系列论文中，[64]我提出了对认为演化过程有可能产生最优性质观点的怀疑。当然，这是未来研究的一个重要领域。

七、结束语

随着中国继续向市场经济的道路上迈进，它将会做出无数决定——创造何种制度，建立什么法律框架，以及采取什么政策。中国的许多政策会受西方目前的争论的影响。这些争论主要受意识形态和分配的政治学所影响。

那些拥有巨额财富的人希望保持并增加这种财富。其他人则希望创造一个更加和谐的社会。前者有时候会认为通过增加国家的财富，会使每一个人从中受益。这个概念被称为涓滴经济学（trickle down）。它没有理论基础，也很少有经验支持。比如近年来，尽管美国 GDP 增加了，而处于中位数的个人收入却下降了：大多数美国人今天的处境与五年以前相比变坏了。在某些情形下，效率和公平之间存在一个权衡。在某些情形下，实行更加平等主义的政策，例如保证每一个人可以得到好的教育和健康，事实上会带来更快的增长。

然而，本文更加关心效率而不是分配问题，尽管二者不能完全分开。现代经济学使我们可以理解无约束的自由市场的局限，证明政府有重要的作

---

64 特别请参见 *Whither Socialism*（1994）；Stiglitz（1974b）的最后一部分；以及“Notes on Evolutionary Economics：Imperfect Capital Markets，Organizational Design，Long-run Efficiency.”1995，编辑重印版本。

用。在中国存在一种危险，在一段政府过度干预的时期以后，会走向另一个极端。这样做不仅会危害短期的经济增长，还会危害长期的经济增长。即便存在某些短期收益，这种增长也不会是可持续的。更重要的是，它会危害创造更加和谐社会的目标。

参考文献

Akerlof, G. (1970). "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism."*Quarterly Journal of Economiscs* 84(3): 488-500.

Arnott, R., B. Greenwald, et al. (1995). "Information and Economic Efficiency." *Information Economics and Policy,Elsevier* 6(1): 77-82.

Arnott, R. and J. E. Stiglitz (1988). "The Basic Analytics of Moral Hazard." *Scandinavian Journal of Economics*90(3): 383-413.

Arnott, R. and J. E. Stiglitz (1991a). "Moral Hazard and Nonmarket Institutions: Dysfunctional Crowding Out of Peer Monitoring?" *American Economic Review* 81 (1): 179-190.

Arnott, R. and J. E. Stiglitz (1991b). Price Equilibrium, Efficiency, and Decentralizability in Insurance Markets. *NBER Working Paper 3642.*

Arrow, K. and G. Debreu (1954). "The Existence of an Equilibrium for a Competitive Economy." *Econometrica* 22(3): 265-290.

Arrow, K. and F. H. Hahn (1971). *General Competitive Analysis* . San Francisco, Holden-Day.

Athar Hussain, Nicholas Stern, et al. (2000). Chinese Reforms from a Comparative Perspective. *Incentives,Organization,and Public Economics:Papers in Honour of sir James Mirrlees*. P. J. Hammond and G. D. Myles, Oxford University Press: 243-277.

Borch, K. (1962). "Equilibrium in a Reinsurance Market." *Econometrica* 30(3): 424–444.

Braverman, A. and J. E. Stiglitz (1982). "Sharecropping and the Interlinking of Agrarian Markets." *American Economic Review* 72(4): 695-715.

Braverman, A. and J. E. Stiglitz (1986). "Cost Sharing Arrangement Under Sharecropping: Moral Hazard, Incentive Flexibility and Risk." *American Journal of Agricultural Economics* 68(3): 642-652.

Brito, D. L., J. H. Hamilton, et al. (1990). "Pareto efficient tax structures." *Oxford Economic Papers* 442: 61-77.

Bruce Greenwald, J. E. Stiglitz, et al. (1984). "Informational Imperfections in the Capital Markets and Macroeconomic Fluctuations." *American Economic Review* 72(4): 194-199.

Dasgupta, P. and J. E. Stiglitz (1977). "Tariffs vs. Quotas as Revenue Raising Devices under Uncertainty." *American Economic Review* 67(5): 975-981.

Debreu, G. (1956). *Theory of Value* . New York, Wiley.

Diamond, P. (1971). "A Model of Price Adjustment." *Journal of Economic Theory* 3(2): 156-168.

Diamond, P. A. (1967). "The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty." *Amcrican Economic Rcvicw* 57(4): 759-776.

Dixit, A. K. and J. E. Stiglitz (1977). "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67(3): 297-308.

Edlin, A. and J. E. Stiglitz (1995). "Discouraging Rivals: Managerial Rent-Seeking and Economic Inefficiencies." *American Economic Review* 85(5): 1301-1312.

Ellerman, D. and J. E. Stiglitz (2000). "New Bridges Across the Chasm: Macro- and Micro-Strategies for Russia and other Transitional Economies." *Zagreb International Review of Economics and Business* 3(1): 41-72.

Ellerman, D. and J. E. Stiglitz (2001). "Not Poles Apart: 'Whither Reform?' and 'Whence Reform?'" *The Journal of Policy Reform* 4(4): 325-338.

G. Ferri, J. E. Stiglitz, et al. (1999). "The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis." *Economic Notes* 28(3): 335-355.

Greenwald, B. and J. E. Stiglitz (1986). "Externalities in Economies with Imperfect Information and Incomplete Markets." *Quarterly Journal of Economics* 101(2): 229-264.

Greenwald, B. and J. E. Stiglitz (1988). "Pareto Inefficiency of Market Economies: Search and Efficiency Wage Models." *American Economic Review* 78(2): 351-355.

Greenwald, B. and J. E. Stiglitz (1990). "Asymmetric Information and the New Theory of the Firm: Financial Constraints and Risk Behavior." *American Economic Review* 80(2): 160–165.

Grossman, S. J. (1977). "The Existence of Futures Markets, Noisy Rational Expectations and Informational Externalities." *Review of Economic Studies* 44(3): 431-449.

Grossman, S. J. and J. E. Stiglitz (1977). "On Value Maximization and Alternative Objectives of the Firm." *The Journal of Finance* 32(2): 389-402.

Grossman, S. J. and J. E. Stiglitz (1980). "Stockholder Unanimity in Making Production and Financial Decisions." *The Quarterly Journal of Economics* 94(3): 543-566.

Grossman, S. J. and J. E. Stiglitz (1987). On the Microeconomics of Technical Progress. *Thechnology Generation in Latin American Manufacturing Industries* J. M. Katz. London, MacMillar Press: 56-77.

Guesnerie, R. and K. Roberts (1984). "Effective Policy Tools and Quantity Controls." *Econometrica* 52(1): 59-86.

Hayek, F. A. (1945). "The Use of Knowledge in Society." *American Economic Review* 35(4): 519-530.

Hoff, K. (1996). "Market Failures and the Distribution of Wealth: A Perspective from the Economics of Information." *Politics and Socicty* 24: 411-432.

Kahneman, D. (2002). Maps of Bounded Rationality. *Nobel Prize Lecture* . Stockholm.

Krugman, P. R. (1980). "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70(5): 950-959.

Lange, O. and F. M. Taylor (1964). *On the Economic Theory of Socialism*. New York, NY, McGraw-Hill.

Leland, H. E. and D. H. Pyle (1977). "Informational Asymmetries, Financial Structure, and Financial Intermediation." *Journal of Finance* 32(2): 371-387.

Lipsey, R. G. and K. Lancaster (1956). "The General Theory of Second Best." *Review of Economic Studies* 24(1): 11-32.

Markowitz, H. (1952). "Portfolio Selection." *Journal of Finance* 7(1): 77-91.

Milgrom, P. and J. Roberts (1992). *Economics,Organization.and Management* . Englewood Cliffs, NJ, Prentice-Hall.

Mirrlees, J. A. (1971). "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38(114): 175-208.
Modigliani, F. and M. H. Miller (1958). "The Cost of Capital, Corporation Finance, and the Theory of Investment." *American Economic Review* 48(3): 261-297.
Myers, S. C. and N. Majluf (1996). Corporate Financing and Investment Decisions When Firms Have Information that Investors Do not Have. *The Theory of Corporate Finance*. M. J. Brennan. Cheltenham, U.K., Elgar. 1: 207–241, previously published in 1984.
Newbery, D. and J. E. Stiglitz (1982). "The Choice of Techniques and the Optimality of Market Equilibrium with Rational Expectations." *Journal of Political Economy* 90(2): 223-246.
Prescott, E. C. and R. M. Townsend (1984). "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard." *Econometrica* 52(1): 21-46.
Radner, R. and J. E. Stiglitz (1984). A Nonconcavity in the Value of Information. *Bayesian Models in Economic Theory*. M. Boyer and R. Khilstrom, Elsevier Science Publications: 33-52.
Ramsey, F. P. (1927). "A Contribution to the Theory of Taxation" *Economic Journal* 37(145): 47-61.
Rey, P. and J. E. Stiglitz (1996). Moral Hazard and Unemployment in Competitive Equilibrium. *Unpublished Manuscript*, University of Toulouse.
Ross, S. A. (1973). "The Economic Theory of Agency: The Principal's Problem." *American Economic Review* 63(2): 134-39.
Rothschild, M. and J. E. Stiglitz (1982). "A Model of Employment Outcomes Illustrating the Effect of the Structure of Information on the Level and Distribution of Income." *Economic Letters* 10: 231-236.
Rothschild, M. and J. E. Stiglitz (1997). "Competition and Insurance Twenty Years Later." *Geneva Papers on Risk and Insurance Theory* 22(2): 73-79.
Sah, R. and J. E. Stiglitz (1985). "Human Fallibility and Economic Organization." *American Economic Review* 75(2): 292-297.
Sah, R. and J. E. Stiglitz (1986). "The Architecture of Economic Systems: Hierarchies and Polyarchies." *American Economic Review* 76(4): 716-727.
Salop, S. and J. E. Stiglitz (1987). Information, Welfare and Product Diversity. *Arrow and the Foundations of the Theory of Economic Policy*. G. Feiwel. London, MacMillan: 328-340.
Samuelson, P. (1947). *Foundations of Economic Analysis*. Cambridge, MA, Harvard University Press.
Samuelson, P. and W. F. Stolper (1941). "Protection and Real Wages." *Review of Economic Studies* 9(1): 58-73.
Sappington, D. and J. E. Stiglitz (1987). "Privatization, Information and Incentives." *Journal of Policy Analysis and Management* 6(4): 567-582.
Shapiro, C. (1983). "Premiums for High Quality Products as Returns to Reputations." *Quarterly Journal of Economics* 98(4): 659-680.
Shapiro, C. and J. E. Stiglitz (1984). "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review* 74(3): 433-444.
Sharpe, W. F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance* 19(4): 425-442.
Shavell, S. (1979). "Risk Sharing and Incentives in the Principal and Agent

Relationship." *The Bell Journal of Economics* 10(1): 55-73.
Shiller, R. J. (2000). *Irrational Exuberance.* Princeton, NJ, Princeton University Press.
Stigler, G. J. (1961). "The Economics of Information." *Journal of Political Economy* 69(3): 213-225.
Stiglitz, J. E. (1969). "A Re-Examination of the Modigliani-Miller Theorem." *American Economic Review* 59(5): 784-793.
Stiglitz, J. E. (1972a). On the Irrelevance of Corporate Financial Policy. *Cowles Foundation Discussion Paper No.339*, Yale University.
Stiglitz, J. E. (1972b). "On the Optimality of the Stock Market Allocation of Investment." *Quarterly Journal of Economics* 86(1): 25-60.
Stiglitz, J. E. (1972c). "Some aspects of the pure theory of corporate finance: Bankruptcies and takeovers." *Bell Journal of Economics* 3(2): 458-482.
Stiglitz, J. E. (1973). "Taxation, Corporate Financial Policy and the Cost of Capital." *Journal of Public Economics* 2(1): 1-34.
tiglitz, J. E. (1974a). "Incentives and Risk Sharing in Sharecropping." *Review of Economic Studies* 41: 219-255.
Stiglitz, J. E. (1974b). Information and Economic Analysis. *Current Economic Problems* J. M. Parkin and A. R. Nobay. Cambridge, Cambridge University Press, Proceedings of the Association of University Teachers of Economics, Manchester, England, April 1974: 27-52.
Stiglitz, J. E. (1975a). Information and Economic Analysis. *Current Economic Problems*. J. M. Parkin and A. R. Nobay. Cambridge, Cambridge University Press: 27-52.
Stiglitz, J. E. (1975b). "The Theory of 'Screening,' Education, and the Distribution of Income." *American Economic Review* 65(3): 283-300.
Stiglitz, J. E. (1982a). Information and Capital Markets. *Financial Economics:Essays in Honor of Paul Cootnet.* W. F. Sharpe and C. Cootner. Englewood Cliffs, NJ, Prentice-Hall 118-58.
Stiglitz, J. E. (1982b). Ownership, Control and Efficient Markets: Some Paradoxes in the Theory of Capital Markets. *Economic Regulation:Essays in Honor of James R.Nelson.* K. D. Boyer.
Stiglitz, J. E. (1982c). "Self-Selection and Pareto Efficient Taxation." *Journal of Public Economics* 17: 213-240.
Stiglitz, J. E. (1982d). "The Inefficiency of the Stock Market Equilibrium." *Review of Economic Studies* 49(2): 241-261.
Stiglitz, J. E. (1984). Information, Screening and Welfare. *Bayesian Models in Economic Theory*. M. Boyer and R. Khilstrom, Elsevier Science Publications: 209-239.
Stiglitz, J. E. (1985a). "Credit Markets and the Control of Capital." *Journal of Money, Credit and Banking* 17(2): 133-152.
Stiglitz, J. E. (1985b). "Equilibrium Wage Distributions." *Economic Journal* 95(379): 595-618.
Stiglitz, J. E. (1986). Toward a More General Theory of Monopolistic Competition. *Prices.Competition and Equilibrium*. M. Peston and R. Quandt, Oxford: Philip Allan/Barnes & Noble Books: 22-69.
Stiglitz, J. E. (1987a). "Competition and the Number of Firms in a Market: Are

Duopolies More Competitive Than Atomistic Markets?" *Journal of Political Economy* 95(5): 1041-1061.

Stiglitz, J. E. (1987b). Design of Labor Contracts: Economics of Incentives and Risk-Sharing. *Incentives,Cooperation and Risk Sharing* . H. Nalbantian. Towota, NJ, Rowman & Allanheld: 47-68.

Stiglitz, J. E. (1987c). Pareto Efficient and Optimal Taxation and the New Welfare Economics. *Handbook on Public Economics.* A. Auerbach and M. Feldstein, North Holland: Elsevier Science Publishers.

Stiglitz, J. E. (1989a). Imperfect Information in the Product Market. *Handbook of Industrial Organization,* Elsevier Science Publishers. 1: 769-847.

Stiglitz, J. E. (1989b). Monopolistic Competition and the Capital Market.*The Economics of Imperfect Competition and Employment-Joan Robinson and Beyond*. G. Feiwel. New York, NY, New York University Press: 485-507.

Stiglitz, J. E. (1989c). Mutual Funds, Capital Structure, and Economic Efficiency. *Theory of Valuation-Frontiers of Modern Financial Theory* S. Bhattacharya and G. Constantinides. Totowa, NJ, Rowman and Littlefield. 1: 342-356.

Stiglitz, J. E. (1989d). "Using tax policy to curb speculative short-term trading " *Journal of Financial Services Research* 3(3/2): 101-115.

Stiglitz, J. E. (1990). "Peer Monitoring and Credit Markets." *World Bank Economic Review* 4(3): 351-366.

Stiglitz, J. E. (1991). The Economic Role of the State: Efficiency and Effectiveness. *Efficiency and Effectiveness in the Public Domain.The Economic Role of the State* . T. P. Hardiman and M. Mulreany, Institute of Public Administration: 37-59.

Stiglitz, J. E. (1993a). Market Socialism and Neoclassical Economics. *Market Socialism: The* Current *Debate* . P. Bardhan and J. Roemer. New York, Oxford University Press: 21-41.

Stiglitz, J. E. (1993b). Perspectives on the Role of Government Risk-Bearing within the Financial Sector. *Government Risk-bearing*. M. Sniderman. Norwell, MA, Kluwer Academic Publishers: 109-130.

Stiglitz, J. E. (1993c). Some Theoretical Aspects of the Privatization: Applications to Eastern Europe. *Privatization Processes in Eastern Europe.* M. Baldassarri, L. Paganetto and E. S. Phelps. Rome, St. Martin's Press: 179-204.

Stiglitz, J. E. (1993d). *The Role of the State in Financial Markets.* Proceeding of the World Bank Conference on Development Economics, Washington, D.C., World Bank.

Stiglitz, J. E. (1993e). Incentives, Organizational Structures, and Contractual Choice in the Reform of Socialist Agriculture. *The Agricultural Transition in Central and Eastern Europe and the Former U.S.S.R* A. Braverman, K. Brooks and C. Csaki, The World Bank: 27-46.

Stiglitz, J. E. (1994). *Whither Socialism*? . Cambridge, MA, MIT Press.

Stiglitz, J. E. (1995). "Notes on Evolutionary Economics: Imperfect Capital Markets, Organizational Design, Long-run Efficiency."

Stiglitz, J. E. (1997). The Role of Government in Economic Development (Keynote Address). *Annual World Bank Conference on Development Economics* 1996 . M. Bruno

and B. Pleskovic, World Bank: 11-23.
Stiglitz, J. E. (2000a). Quis custodiet ipsos custodes? Corporate Governance Failures in the Transition. *Governance,Equity and Global Markets,Proceedings from the Annual Bank Conference on Development Economics in Europe,June* 1999.P.-A. Muet and J. E. Stiglitz. Paris, Conseil d'Analyse Economique: 51-84.
Stiglitz, J. E. (2000b). Whither Reform? Ten Years of Transition. *Annual World Bank Conference on Economic Development*. B. Pleskovic and J. E. Stiglitz. Washington, World Bank: 27-56.
Stiglitz, J. E. (2001a). "Principles of Financial Regulation: A Dynamic Approach." *The World Bank Observer* 16(1): 1-18.
Stiglitz, J. E. (2001b). Robust Financial Restraint. *Financial Liberalization:How Far, How Fast*?. G. Caprio, P. Honohan and J. Stiglitz. Cambridge, UK, Cambridge University Press: 31-63.
Stiglitz, J. E. (2002). "Information and the Change in the Paradigm in Economics, abbreviated Nobel Lecture." *American Economic Review* 92(3): 460-501.
Stiglitz, J. E. (2003a). On Liberty, the Right to Know, and Public Discourse: The Role of Transparency in Public Life. *Globalizing Rights* M. J. Gibney, Oxford University Press: 115-156.
Stiglitz, J. E. (2003b). *Roaring Nineties* New York, NY, W W Norton.
Stiglitz, J. E. (2004). Reflections on the State of the Theory of Monopolistic Competition. *The monopolistic competition revolution in retrospect*. S. Brakman and B. J. Heijdra. Cambridge, Cambridge University Press.
Stiglitz J. E. (2006). *Marking Globalization Work*,WW Norton.
Stiglitz, J. E. (forthcoming). Information and Political Processes. *The 2004 Wildasky Lecture* Berkelely University Press.
Stiglitz, J. E. and A. Weiss (1983). "Incentive Effects of Termination: Applications to the Credit and Labor Markets." *American Economic Review* 73(5): 912-927.
Stiglitz, J. E. and J. Yun (2002). Integration of Unemployment Insurance with Retirement Insurance. *NBER Working Paper No.*9199.
Thaler, R. H. (1990). "Anomalies: Saving, Fungibility, and Mental Accounts." *The Journal of Economic Perspectives* 4(1): 193-205.
Tversky, A. and D. Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases" *Science* 185(4157): 1124-1131.

Bibliography: J. E. Stiglitz and Welfare Economics

I. Basic articles

"Externalities in Economies with Imperfect Information and Incomplete Markets," with B. Greenwald, *Quarterly Journal of Economics*, May 1986, pp. 229-264. (volume 1)
"Pareto Inefficiency of Market Economies: Search and Efficiency Wage Models," with B. Greenwald, *American Economic Review*, 78(2), May 1988, pp. 351-355.
"Moral Hazard and Non-Market Institutions: Dysfunctional Crowding Out or Peer Monitoring," with R. Arnott, *American Economic Review*, 81(1), March 1991, pp.

179-190.

"Information and Economic Efficiency," with R. Arnott and B. Greenwald, *Information Economics and Policy*, 6(1), March 1994, pp. 77-88.

"Privatization, Information and Incentives," with D. Sappington, *Journal of Policy Analysis and Management*, 6(4), 1987, pp. 567-582 reprinted in E. Baily and J. Hower (eds.), *The Political Economy of Privatization and Deregulation*, Edward Elgar, 1993.

II. Interpretation

The Invisible Hand and Modern Welfare Economics," In *Information Strategy and Public Policy*, D. Vines and A. Stevenson (eds.), Oxford: Basil Blackwell, 1991, pp. 12-50. (Stevenson Lecture given at Glasgow University, December 1988, NBER Working Paper 3641.)

"Market Socialism and Neoclassical Economics," In *Market Socialism: The Current Debate*, P. Bardhan and J. Roemer (eds.), New York: Oxford University Press, 1993, pp. 21–41.

III. Applications

Uncertainty

"On the Optimality of the Stock Market Allocation of Investment," *Quarterly Journal of Economics*, 86(1), February 1972, pp. 25-60.

"The Inefficiency of the Stock Market Equilibrium," *Review of Economic Studies*, XLIX, April 1982, pp. 241-2611.

"The Choice of Techniques and the Optimality of Market Equilibrium with Rational Expectations," with D. Newbery, *Journal of Political Economy*, 90(2), April 1982, pp. 223-246.

"Pareto Inferior Trade," with D. Newbery, Review of Economic Studies, 51(1), January 1984, pp. 1-12.

"Mutual Funds, Capital Structure, and Economic Efficiency," In *Theory of Valuation - Frontiers of Modern Financial Theory, Vol. 1*, S. Bhattacharya and G. Constantinides (eds.), Totowa, NJ: Rowman and Littlefield, 1989, pp. 342-356.

natural resources

"The Efficiency of Market Prices in Long Run Allocations in the Oil Industry," In *Studies in Energy Tax Policy*, G. Brannon (ed.), Cambridge: Ballinger Publishing, 1975, pp. 55-99.

information

"Equilibrium Unemployment as a Worker Discipline Device," with Carl Shapiro, *American Economic Review*, 74(3), June 1984, pp. 433-444.

"The Welfare Economics of Moral Hazard," with R. Arnott, in *Risk Information and Insurance: Essays in the Memory of Karl H. Borch*, H. Louberge (ed.), Norwell: Kluwer Academic Publishers, 1990, pp. 91-122.

"Information, Screening and Welfare," In *Bayesian Models in Economic Theory*, Marcel Boyer and Richard Khilstrom (eds.), Elsevier Science Publications, 1984, pp. 209-239.

"Labor Turnover, Wage Structure & Moral Hazard: The Inefficiency of Competitive Markets," with R. Arnott, *Journal of Labor Economics*, 3(4), October 1985, pp. 434-462.

"On the Impossibility of Informationally Efficient Markets," with S. Grossman, *American Economic Review*, 70(3), June 1980, pp. 393-408.

"Information, Welfare and Product Diversity," with S. Salop, In *Arrow and the Foundations of the Theory of Economic Policy*, G. Feiwel (ed.), London: MacMillan, 1987, pp. 328-340.

Discouraging Rivals: Managerial Rent-Seeking and Economic Inefficiencies," with A. Edlin, *American Economic Review*, 85(5), December 1995, pp. 1301–12. (Also NBER Working Paper 4145, 1992.)

"Remarks on Inequality, Agency Costs, and Economic Efficiency," Paper prepared for a workshop in "Economic Theories of Inequality," Stanford Institute for Theoretical Economics, Stanford University, March 11–13, 1993.

"Using Tax Policy to Curb Speculative Short-Term Trading," *Journal of Financial Services Research*, 3(2/3), December 1989, pp. 101-115.

"Information and Capital Markets," In *Financial Economics: Essays in Honor of Paul Cootner*, William F. Sharpe and Cathryn Cootner (eds.), Prentice Hall, New Jersey, 1982, pp. 118-158. (Also NBER Working Paper 678.)

"Information Externalities in Oil and Gas Leasing," with J. Leitzinger, *Contemporary Policy Issues*, 5, March 1984, pp. 44-57. (Paper presented at the Western Economic Association Meetings, July 1983.)

"The Theory of Screening, Education and the Distribution of Income," *American Economic Review*, 65(3), June 1975, pp. 283-300.

"Information and Economic Analysis," In J.M. Parkin and A.R. Nobay, eds. *Current Economic Problems* (Cambridge: Cambridge University Press), pp. 27-52.

IV. Efficiency and Competition

"Monopolistic Competition and Optimum Product Diversity," with A. Dixit, *American Economic Review*, 67(3), June 1977, pp. 297-308.

"Pareto Optimality and Competition," *Journal of Finance*, 36(2), May 1981, pp. 235-251.

"Potential Competition May Reduce Welfare," *American Economic Review*, 71(2), May 1981, pp.184-189. (Papers and Proceedings of the AEA meetings in Denver, Colo., September 1980.)

"Technological Change, Sunk Costs, and Competition," *Brookings Papers on Economic Activity*, 3, 1987. (Special issue of *Microeconomics*, M.N. Baily and C. Winston (eds.), 1988, pp. 883-947.)

"Potential Competition, Actual Competition and Economic Welfare," with P. Dasgupta, European Economic Review, 32, May 1988, pp. 569-577.

"Learning by Doing, Market Structure, and Industrial and Trade Policies," with P. Dasgupta, Oxford Economic Papers, 40(2), 1988, pp. 246-268.

封面
书名
版权
目录
第三卷导言
风险原则
递增的风险：一个定义
递增的风险：经济意义
投资者偏好和资产回报的结构与投资组合配置的分离性：对共同基金纯理论的贡献
不确定性下的厂商理论
关于公司财务纯理论的几个问题：破产与收购
股市投资配置的最优性
价值最大化与企业的可选择目标
关于公司财务政策的无关性
有风险的市场均衡
论股票市场均衡的无效性
帕累托次优贸易
资本市场自由化、全球化与国际货币基金组织
垄断竞争理论
垄断竞争与最优产品多样性
创新理论
关于技术变革的新观点
产业结构与创新活动的本质
不确定性、产业结构与研发速度
技术变迁、沉没成本与竞争
干中学、市场结构与产业政策和贸易政策
福利经济学
福利经济学